บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยศึกษาภายใต้การจำลองข้อมูลด้วยการจัดกระทำปัจจัย 4 ตัว คือ (1) ลักษณะของข้อสอบ 9 ลักษณะ (a ต่ำกับ b ต่ำ, a ต่ำกับ b ปานกลาง, a ต่ำกับ b สูง, a ปานกลางกับ b ต่ำ, a ปานกลางกับ b ปานกลาง, a ปานกลางกับ b สูง, a สูงกับ b ต่ำ, a สูงกับ b ปานกลาง และ a สูงกับ b สูง) (2) ความยาวของแบบสอบ 2 ระดับ (30 ข้อ และ 60 ข้อ) (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ (5%, 10% และ 20%) และ (4) ขนาดกลุ่มตัวอย่าง 6 ระดับ (จำนวนผู้สอบกลุ่มอ้างอิงต่อจำนวนผู้สอบกลุ่มเปรียบเทียบเท่ากับ 250 คนต่อ 250 คน, 500 คนต่อ 250 คน, 500 คนต่อ 500 คน, 1000 คนต่อ 250 คน, 1000 คนต่อ 500 คน และ 1000 คนต่อ 1000 คน)

ผู้วิจัยใช้โปรแกรม IRTDATA (Johanson, 1992) จำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ (3PLM-c) โดยจำลองข้อมูล 9 เมทริกซ์ ในแต่ละเมทริกซ์มีข้อสอบ 90 ข้อ และผู้สอบ 2,000 คน แล้วนำข้อมูลที่จำลองไปจัดกระทำตามปัจจัยที่ศึกษา 4 ตัว ในการจัดกระทำข้อมูลดังกล่าวจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบภายใต้ทฤษฎีการตอบสนองข้อสอบ ซึ่งมี 4 ขั้นตอน กล่าวคือ ***ขั้นตอนแรก*** ใช้โปรแกรม BILOG version 3.04 (Mislevy and Bock, 1990) ประมาณค่าพารามิเตอร์ของข้อสอบภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ ***ขั้นตอนที่สอง*** ใช้โปรแกรม EQUATE version 2.0 (Baker, 1993) ปรับเทียบสเกลพารามิเตอร์ของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบให้อยู่บนเมทริกซ์เดียวกันโดยใช้วิธีโค้งลักษณะแบบสอบ (test characteristic curve; TCC) ของ Stocking และ Lord (1983 cited in Kim and Cohen, 1992a) ภายใต้โมเดลโลจิสติกแบบ 2 พารามิเตอร์ ***ขั้นตอนที่สาม*** ใช้โปรแกรม IRTDIF version 1.0 (Kim and Cohen, 1992b) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปโดยใช้วิธีการวัดพื้นที่ในช่วงเปิดชนิดไม่คิดเครื่องหมายของ Raju (1990) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ ***ขั้นตอนสุดท้าย*** ใช้โปรแกรม SPSS for windows version 7.52 สุ่มข้อมูลตามปัจจัยที่ศึกษา 4 ตัว รวมข้อมูลที่จัดกระทำทั้งหมด 324 เงื่อนไข (9 × 2 × 3 × 6) ต่อจากนั้นจึงนำข้อมูลที่จัดกระทำดังกล่าวไปวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ

อเนกรูป 4 วิธี คือ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอย-โลจิสติก โดยสองวิธีแรกใช้โปรแกรม SIBTEST ของ Stout และ Roussos version 1.1 (1992) วิธีต่อมาใช้โปรแกรม MHDIF version 1.0 (Fidalgo, 1995) และวิธีสุดท้ายใช้โปรแกรม SPSS/PC$^+$ version 4.01

ในการวิเคราะห์อำนาจการทดสอบ (ร้อยละของจำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกัน) และอัตราความคลาดเคลื่อนประเภทที่ 1 (ร้อยละของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกจะใช้วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมายของ Raju เป็นเกณฑ์ในการเปรียบเทียบ กล่าวคือ ถ้าวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกระบุข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของ Raju แสดงว่า วิธีดังกล่าวสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปได้ถูกต้อง และถ้าวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกระบุข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปไม่ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของ Raju แสดงว่า วิธีดังกล่าวตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปไม่ถูกต้อง ส่วนการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษา 4 ตัว ผู้วิจัยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบมีนัยสำคัญทางสถิติจะใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ทดสอบในแต่ละระดับของตัวแปรแล้วเปรียบเทียบความแตกต่างเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ยกเว้นการเปรียบเทียบภายใต้ความยาวของแบบสอบ 2 ระดับจะใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (independent sample) โดยทดสอบใน 3 กรณีต่อไปนี้ (1) ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน (2) ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ (3) ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษาต่างระดับกัน

สำหรับการนำเสนอผลการวิเคราะห์ข้อมูล ผู้วิจัยแบ่งออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

ตอนที่ 2 ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

ตอนที่ 3 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

เพื่อความสะดวกในการนำเสนอผลการวิเคราะห์ข้อมูล ผู้วิจัยกำหนดสัญลักษณ์ ดังนี้

$\overline{X}$ หมายถึง ค่าเฉลี่ยเลขคณิต

SD หมายถึง ส่วนเบี่ยงเบนมาตรฐาน

TI หมายถึง ลักษณะของข้อสอบ

a_L หมายถึง ข้อสอบที่มีค่าอำนาจจำแนกต่ำ

a_M หมายถึง ข้อสอบที่มีค่าอำนาจจำแนกปานกลาง

a_H หมายถึง ข้อสอบที่มีค่าอำนาจจำแนกสูง

b_L หมายถึง ข้อสอบที่มีค่าความยากต่ำ

b_M หมายถึง ข้อสอบที่มีค่าความยากปานกลาง

b_H หมายถึง ข้อสอบที่มีค่าความยากสูง

TL หมายถึง ความยาวของแบบสอบ

% DIF หมายถึง สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

N_R หมายถึง ขนาดผู้สอบกลุ่มอ้างอิง

N_F หมายถึง ขนาดผู้สอบกลุ่มเปรียบเทียบ

MSIB หมายถึง วิธีซิปเทสท์ปรับใหม่

SIB หมายถึง วิธีซิปเทสท์

MH หมายถึง วิธีแมนเทล-แฮนส์เซล

LR หมายถึง วิธีการถดถอยโลจิสติก

P หมายถึง อำนาจการทดสอบ

E_1 หมายถึง อัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 1 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

ผู้วิจัยนำข้อมูลที่จำลองมาจัดกระทำตามปัจจัยที่ศึกษา 4 ตัว คือ ลักษณะของข้อสอบ 9 ลักษณะ ความยาวของแบบสอบ 2 ระดับ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ และขนาดกลุ่มตัวอย่าง 6 ระดับ รวมทั้งหมด 324 เงื่อนไข (9 × 2 × 3 × 6) แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป 4 วิธี คือ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ผลการวิเคราะห์จำแนกตามปัจจัยที่ศึกษาดังตารางที่ 6

ตารางที่ 6 ค่าเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของร้อยละของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ซึ่งตรวจสอบด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	ร้อยละของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
TI								
a_L , b_L	23.24	14.66	6.57	5.72	9.54	5.64	17.04	8.23
a_L , b_M	22.22	9.73	10.19	4.84	11.39	7.19	18.43	6.92
a_L , b_H	17.41	7.35	7.31	5.33	12.50	7.01	15.33	8.01
a_M , b_L	16.11	9.61	7.22	5.13	11.57	28.20	13.52	7.17
a_M , b_M	16.11	7.19	8.80	4.30	9.07	5.50	15.00	8.49
a_M , b_H	12.50	6.14	7.59	4.03	7.41	4.30	14.44	6.17
a_H , b_L	12.31	6.98	5.65	5.27	5.37	5.06	11.39	7.01
a_H , b_M	15.09	6.25	7.50	5.06	8.89	5.58	14.63	7.05
a_H , b_H	16.20	13.31	8.24	5.31	9.54	5.97	13.89	7.28

ตารางที่ 6 (ต่อ)

ปัจจัย	ร้อยละของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
TL								
30 ข้อ	12.90	7.53	5.70	4.10	7.35	13.90	10.56	4.83
60 ข้อ	20.70	10.69	9.65	5.28	11.60	6.22	19.12	7.36
%DIF								
5 %	14.57	7.29	6.94	4.63	8.89	6.28	13.67	7.07
10 %	14.54	6.96	7.93	5.07	9.04	5.93	14.51	6.98
20 %	16.80	10.03	7.67	5.11	9.48	10.96	14.85	7.55
$N_R : N_F$								
250 : 250	15.31	7.94	7.47	4.93	8.77	5.47	16.05	7.41
500 : 250	16.85	8.50	7.78	5.79	8.27	6.30	14.15	8.94
500 : 500	17.04	8.68	7.96	5.16	11.60	23.50	15.49	7.36
1000 : 250	17.04	10.39	8.46	5.74	8.21	5.60	14.38	7.70
1000 : 500	16.48	10.12	6.98	4.59	10.74	6.03	13.77	6.73
1000 : 1000	18.09	13.68	7.41	4.42	9.26	5.74	15.25	7.05

จากตารางที่ 6 ค่าเฉลี่ยเลขคณิต ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของร้อยละของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป คำนวณมาจากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากวิธีการตรวจสอบ 4 วิธีภายใต้ข้อมูลที่จัดกระทำทั้งหมด 324 เงื่อนไข เมื่อจำแนกตามลักษณะของข้อสอบจะมีข้อสอบที่แตกต่างกันทางด้านค่าอำนาจจำแนก (a) และค่าความยาก (b) 9 ลักษณะ (1 ลักษณะต่อข้อมูล 36 เงื่อนไข) เมื่อจำแนกตามความยาวของแบบสอบจะมีจำนวนข้อสอบที่แตกต่างกัน 2 ระดับ (1 ระดับต่อข้อมูล 162 เงื่อนไข) เมื่อจำแนกตามสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันจะมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ 3 ระดับ (1 ระดับต่อข้อมูล

108 เงื่อนไข) เมื่อจำแนกตามขนาดกลุ่มตัวอย่างจะมีข้อสอบที่มีผู้สอบ 6 ระดับ (1 ระดับต่อข้อมูล 54 เงื่อนไข)

เมื่อพิจารณาปัจจัยลักษณะของข้อสอบ (TI) พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_L โดยเฉลี่ยแล้ววิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 23.24 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_L สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 12.31 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_M โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 10.19 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_L สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 5.65 ส่วนของวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_H สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ประมาณร้อยละ 12.50 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_L สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 5.37 ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_M โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 18.43 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_L สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 11.39

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกันทั้ง 9 ลักษณะ พบว่า ภายใต้เกือบทุกลักษณะของข้อสอบ วิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ส่วนวิธีซิปเทสท์สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด

เมื่อพิจารณาปัจจัยความยาวของแบบสอบ (TL) พบว่า ภายใต้แบบสอบที่มีความยาว 30 ข้อ โดยเฉลี่ยแล้ววิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันประมาณร้อยละ 12.90 และภายใต้แบบสอบที่มีความยาว 60 ข้อ สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยประมาณร้อยละ 20.70 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้แบบสอบที่มีความยาว 30 ข้อ โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันประมาณร้อยละ 5.70 และภายใต้แบบสอบที่มีความยาว 60 ข้อ สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยประมาณร้อยละ 9.65 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้แบบสอบที่มีความยาว 30 ข้อ โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันประมาณร้อยละ 7.35 และภายใต้แบบสอบที่มีความยาว 60 ข้อ โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของ

ข้อสอบที่ทำหน้าที่ต่างกันประมาณร้อยละ 11.60 สำหรับวิธีการถดถอยโลจิสติก พบว่า ภายใต้แบบสอบที่มีความยาว 30 ข้อ โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันประมาณร้อยละ 10.56 และภายใต้แบบสอบที่มีความยาว 60 ข้อ สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยประมาณร้อยละ 19.12

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้แบบสอบที่มีความยาวต่างกันทั้ง 2 ระดับ พบว่า โดยเฉลี่ยแล้ววิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ส่วนวิธีซิปเทสท์สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด

เมื่อพิจารณาปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% โดยเฉลี่ยแล้ววิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 16.80 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ 10% สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 14.54 ส่วนวิธีซิปเทสท์ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ประมาณร้อยละ 7.93 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 6.94 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 9.48 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด ประมาณร้อยละ 8.89 ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ 20% โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 14.85 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 13.67

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันทั้ง 3 ระดับ พบว่า วิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ในขณะที่วิธีซิปเทสท์สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด

เมื่อพิจารณาปัจจัยขนาดกลุ่มตัวอย่าง (N_R : N_F) พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 วิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ประมาณร้อยละ 18.09 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 15.31 ส่วนวิธีซิปเทสท์ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 250 โดยเฉลี่ยแล้วสามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ประมาณร้อยละ 8.46 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 500 สามารถตรวจพบโดยเฉลี่ยน้อยที่สุด ประมาณร้อยละ 6.98 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 500 สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ประมาณร้อยละ 11.60 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 250 สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 8.21 ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ประมาณร้อยละ 16.05 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 500 สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด ประมาณร้อยละ 13.77

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกันทั้ง 6 ระดับ พบว่า ภายใต้เกือบทุกขนาดกลุ่มตัวอย่าง วิธีซิปเทสท์ปรับใหม่สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยมากที่สุด ส่วนวิธีซิปเทสท์สามารถตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันเฉลี่ยน้อยที่สุด

ตอนที่ 2 ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

ผู้วิจัยนำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ไปวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งใช้วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมายของ Raju เป็นเกณฑ์ในการเปรียบเทียบ โดยการวิเคราะห์อำนาจการทดสอบ (ร้อยละของจำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกัน) คำนวณได้จากจำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกันซึ่งตรวจสอบด้วยวิธีที่ศึกษาต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบซึ่ง

ตรวจสอบด้วยวิธีการวัดพื้นที่ของ Raju ส่วนการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 (ร้อยละของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) คำนวณได้จากจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันซึ่งตรวจสอบด้วยวิธีที่ศึกษาต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบที่ตรวจสอบด้วยวิธีการวัดพื้นที่ของ Raju ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในรูปค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานแสดงดังตารางที่ 7 และ 8 สำหรับรายละเอียดของการวิเคราะห์ในแต่ละเงื่อนไขแสดงดังตารางที่ 102 ในภาคผนวก ฉ หน้า 350 – 359

ตารางที่ 7 ค่าเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	อำนาจการทดสอบ							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
TI								
a_L , b_L	26.36	27.62	18.03	22.78	19.17	18.76	24.33	23.30
a_L , b_M	53.22	28.43	12.69	19.25	19.22	19.61	39.75	25.66
a_L , b_H	27.53	32.30	22.89	18.58	18.53	20.60	22.39	24.46
a_M , b_L	25.11	23.72	6.72	10.50	5.78	13.65	15.54	18.60
a_M , b_M	45.81	36.40	25.17	19.45	29.14	31.64	49.28	33.11
a_M , b_H	10.86	13.55	5.11	8.11	5.53	11.02	30.69	22.29
a_H , b_L	8.08	13.25	1.83	6.29	15.97	23.99	19.48	20.41
a_H , b_M	37.64	26.66	16.61	17.17	26.08	22.35	49.75	26.80
a_H , b_H	23.17	25.29	10.92	14.91	8.61	14.21	19.69	16.36

ตารางที่ 7 (ต่อ)

ปัจจัย	อำนาจการทดสอบ							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
TL								
30 ข้อ	28.17	30.87	13.05	18.69	16.77	25.28	31.38	27.27
60 ข้อ	29.12	28.14	13.61	16.59	16.12	17.39	28.82	26.03
%DIF								
5 %	27.29	34.98	14.94	22.55	20.94	28.13	38.24	32.13
10 %	30.08	28.87	14.19	16.97	17.13	21.28	31.00	27.21
20 %	28.56	23.77	10.85	11.58	11.28	11.19	21.06	14.85
$N_R : N_F$								
250 : 250	20.00	19.56	9.09	14.66	10.80	16.22	24.38	20.32
500 : 250	21.09	24.21	10.17	14.58	12.72	16.48	22.63	21.64
500 : 500	30.41	29.14	12.63	18.12	13.09	17.83	27.93	25.08
1000 : 250	24.28	26.77	12.65	17.94	13.41	20.10	24.37	20.02
1000 : 500	33.94	34.37	14.17	19.20	22.54	24.90	34.54	32.98
1000 : 1000	42.13	34.86	21.28	18.90	26.13	27.98	46.76	29.96

จากตารางที่ 7 ค่าเฉลี่ยของอำนาจการทดสอบ (ร้อยละของจำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกัน) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยที่ศึกษา ปรากฏผลดังนี้

เมื่อพิจารณาปัจจัยลักษณะของข้อสอบ (TI) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีซิปเทสท์ปรับใหม่ พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 53.22 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าเฉลี่ยต่ำสุด ประมาณ

ร้อยละ 8.08 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 25.17 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 1.83 ส่วนวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 29.14 แต่ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 5.53 สำหรับวิธีการถดถอยโลจิสติก ปรากฏว่า ภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 49.75 แต่ภายใต้ลักษณะของข้อสอบ a_M, b_L มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 15.54

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ลักษณะของข้อสอบต่างกันทั้ง 9 ลักษณะ พบว่า ภายใต้ทุกลักษณะของข้อสอบที่มีค่าความยากปานกลาง (a_L, b_M; a_M, b_M และ a_H, b_M) วิธีซิปเทสท์ปรับใหม่ วิธีการถดถอยโลจิสติก และวิธีแมนเทล-แฮนส์เซล มีค่าเฉลี่ยของอำนาจการทดสอบสูงกว่าภายใต้ทุกลักษณะของข้อสอบที่มีค่าความยากต่ำและสูง (a_L, b_L; a_L, b_H; a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H)

เมื่อพิจารณาปัจจัยความยาวของแบบสอบ (TL) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยของอำนาจการทดสอบประมาณร้อยละ 28.17 และภายใต้ความยาวของแบบสอบ 60 ข้อมีค่าเฉลี่ยประมาณร้อยละ 29.12 ส่วนวิธีซิปเทสท์ พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยของอำนาจการทดสอบประมาณร้อยละ 13.05 และภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยประมาณร้อยละ 13.61 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยของอำนาจการทดสอบประมาณร้อยละ 16.77 แต่ภายใต้ความยาวของแบบสอบ 60 ข้อมีค่าเฉลี่ยประมาณร้อยละ 16.12 ส่วนวิธีการถดถอยโลจิสติก พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยของอำนาจการทดสอบประมาณร้อยละ 31.38 และภายใต้ความยาวของแบบสอบ 60 ข้อมีค่าเฉลี่ยประมาณร้อยละ 28.82

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ความยาวของแบบสอบต่างกันทั้ง 2 ระดับ ปรากฏว่า ภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ ค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าใกล้เคียงกัน ส่วนวิธีการถดถอยโลจิสติก ปรากฏว่า ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยของอำนาจการทดสอบสูงกว่าภายใต้ความยาวของแบบสอบ 60 ข้อ

เมื่อพิจารณาปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 30.08 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 27.29 ส่วนวิธีซิปเทสท์ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 14.94 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 10.85 ส่วนการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 20.94 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 11.28 สำหรับการตรวจสอบด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 38.24 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน20% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 21.06

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 ระดับ ปรากฏว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีค่าเพิ่มขึ้นในทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (5%, 10% และ 20%) แล้วค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าลดลง

เมื่อพิจารณาปัจจัยขนาดกลุ่มตัวอย่าง (N_R : N_F) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 38.26 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 20.00 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 21.28 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 9.09 สำหรับการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 26.13 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 10.80 ส่วนวิธีการถดถอยโลจิสติก พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 40.59 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 500 : 250 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 22.63

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกันทั้ง 6 ระดับ ปรากฏว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าสูงสุดในเงื่อนไขของขนาดกลุ่มตัวอย่างขนาดใหญ่สุด ($N_R : N_F = 1000 : 1000$)

ตารางที่ 8 ค่าเฉลี่ย ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	อัตราความคลาดเคลื่อนประเภทที่ 1							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
TI								
a_L , b_L	9.50	7.64	2.94	3.21	4.53	3.20	9.94	4.57
a_L , b_M	10.64	5.14	5.00	2.97	6.31	3.72	9.25	3.96
a_L , b_H	10.78	5.72	4.28	3.79	7.39	3.57	9.33	5.09
a_M , b_L	9.28	4.95	4.39	2.90	4.92	4.02	8.06	4.07
a_M , b_M	7.36	4.52	4.17	2.90	3.69	3.03	6.19	4.25
a_M , b_H	8.64	4.87	5.64	4.41	5.56	3.93	8.17	3.98
a_H , b_L	7.92	4.42	3.89	3.45	2.31	3.03	6.36	3.63
a_H , b_M	7.56	4.36	3.58	2.67	3.75	2.99	5.36	4.23
a_H , b_H	8.25	4.65	4.89	3.38	5.64	4.46	7.92	4.36
TL								
30 ข้อ	9.62	6.03	4.84	4.06	5.07	4.56	8.04	5.06
60 ข้อ	8.14	4.36	3.78	2.43	4.73	2.92	7.64	3.76

ตารางที่ 8 (ต่อ)

ปัจจัย	อัตราความคลาดเคลื่อนประเภทที่ 1							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
%DIF								
5 %	9.02	5.23	4.35	3.21	5.06	3.84	7.61	4.20
10 %	7.87	4.56	4.64	3.81	5.19	3.71	7.63	4.44
20 %	9.75	5.91	3.94	3.07	4.44	3.92	8.29	4.73
$N_R : N_F$								
250 : 250	8.69	4.27	4.69	3.17	5.52	4.02	9.28	4.65
500 : 250	10.00	5.40	4.61	3.57	4.69	3.72	8.09	4.85
500 : 500	9.04	5.44	4.56	3.68	4.67	4.40	8.63	4.43
1000 : 250	10.04	5.82	5.02	3.62	4.74	3.42	7.94	4.82
1000 : 500	8.24	5.16	3.59	2.76	5.61	3.95	6.93	4.01
1000 : 1000	7.28	5.29	3.39	3.22	4.17	3.30	6.19	3.26

จากตารางที่ 8 ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 (ร้อยละของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยที่ศึกษา ปรากฏผลดังนี้

เมื่อพิจารณาปัจจัยลักษณะของข้อสอบ (TI) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ ปรากฏว่า ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 10.78 แต่ภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 7.36 ส่วนวิธีซิปเทสท์ พบว่า ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.64 แต่ภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 2.94 สำหรับวิธีแมนเทล-แฮนส์เซล

พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 7.39 และภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 2.31 ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 9.94 แต่ภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 5.36

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกันทั้ง 9 ระดับ ปรากฏว่า ภายใต้ทุกลักษณะของข้อสอบที่มีค่าอำนาจจำแนกต่ำ (a_L, b_L; a_L, b_M และ a_L, b_H) ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าภายใต้ลักษณะของข้อสอบที่มีค่าอำนาจจำแนกปานกลางและสูง (a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H)

เมื่อพิจารณาปัจจัยความยาวของแบบสอบ (TL) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ ปรากฏว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 9.62 และ 8.14 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 4.84 และ 3.78 ตามลำดับ สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 5.07 และ 4.73 ตามลำดับ ส่วนวิธีการถดถอยโลจิสติก พบว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 8.04 และ 7.64 ตามลำดับ

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ความยาวของแบบสอบต่างกันทั้ง 2 ระดับ ปรากฏว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติก ภายใต้แบบสอบ 30 ข้อและ 60 ข้อมีค่าใกล้เคียงกัน ส่วนวิธีซิปเทสท์-ปรับใหม่และวิธีแมนเทล-แฮนส์เซล ปรากฏว่า ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ความยาวของแบบสอบ 60 ข้อ

เมื่อพิจารณาปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ปรับใหม่ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 9.75 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 7.87 ส่วนวิธีซิปเทสท์ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าเฉลี่ยของอัตราความ-

คลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 4.64 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 3.94 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.19 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 4.44 ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 8.29 แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 7.61

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 ระดับ พบว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนเพิ่มขึ้นในทุกเงื่อนไข (5%, 10% และ 20%) แล้วค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น

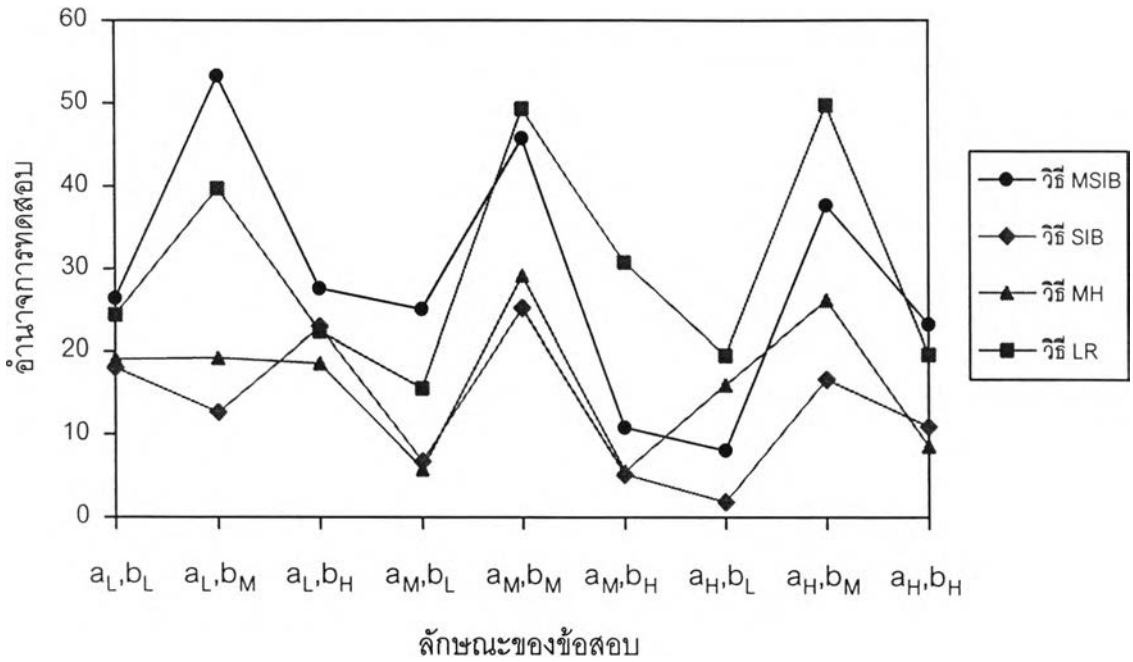
เมื่อพิจารณาปัจจัยขนาดกลุ่มตัวอย่าง (N_R : N_F) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีซิปเทสท์ปรับใหม่ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 250 มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 10.04 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 7.28 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 250 มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.02 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 3.39 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 500 มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.61 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 4.17 ส่วนวิธีการถดถอยโลจิสติก พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 9.28 แต่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าเฉลี่ยต่ำสุด ประมาณร้อยละ 6.19

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกันทั้ง 6 ระดับ ปรากฏว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าต่ำสุดในเงื่อนไขของขนาดกลุ่มตัวอย่างขนาดใหญ่สุด (N_R : N_F = 1000 : 1000)

ตอนที่ 3 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

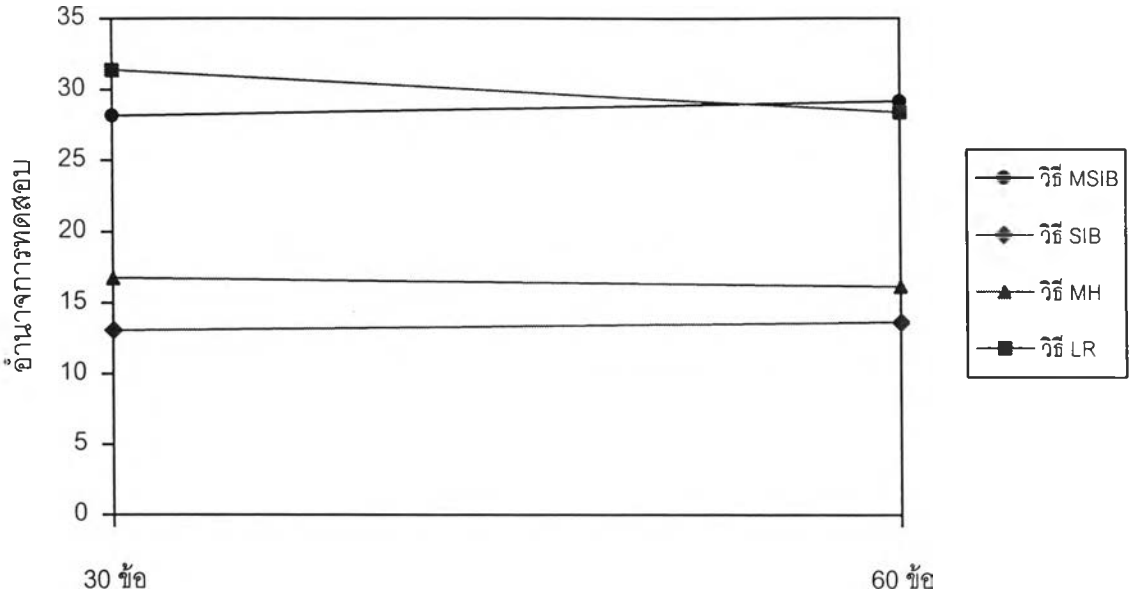
1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา

ก่อนที่จะนำผลการวิเคราะห์อำนาจการทดสอบจากตารางที่ 7 ไปทดสอบทางสถิติ ผู้วิจัยนำผลการวิเคราะห์ดังกล่าวไปสร้างกราฟ เพื่อพิจารณาผลกระทบอย่างคร่าว ๆ ของตัวแปรอิสระที่มีต่อตัวแปรตาม ตัวแปรอิสระมี 2 ตัวหลัก คือ (1) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี ได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล วิธีการถดถอยโลจิสติก และ (2) ปัจจัยที่ศึกษา 4 ตัว ได้แก่ ลักษณะของข้อสอบ 9 ลักษณะ (a_L, b_L; a_L, b_M; a_L, b_H; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H) ความยาวของแบบสอบ 2 ระดับ (30 ข้อ และ 60 ข้อ) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ (5%, 10% และ 20%) และขนาดกลุ่มตัวอย่าง 6 ระดับ (N_R : N_F = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000) ส่วนตัวแปรตาม คือ อำนาจการทดสอบ กราฟดังกล่าวแสดงดังภาพที่ 4, 5, 6 และ 7



ภาพที่ 4 กราฟแสดงการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ

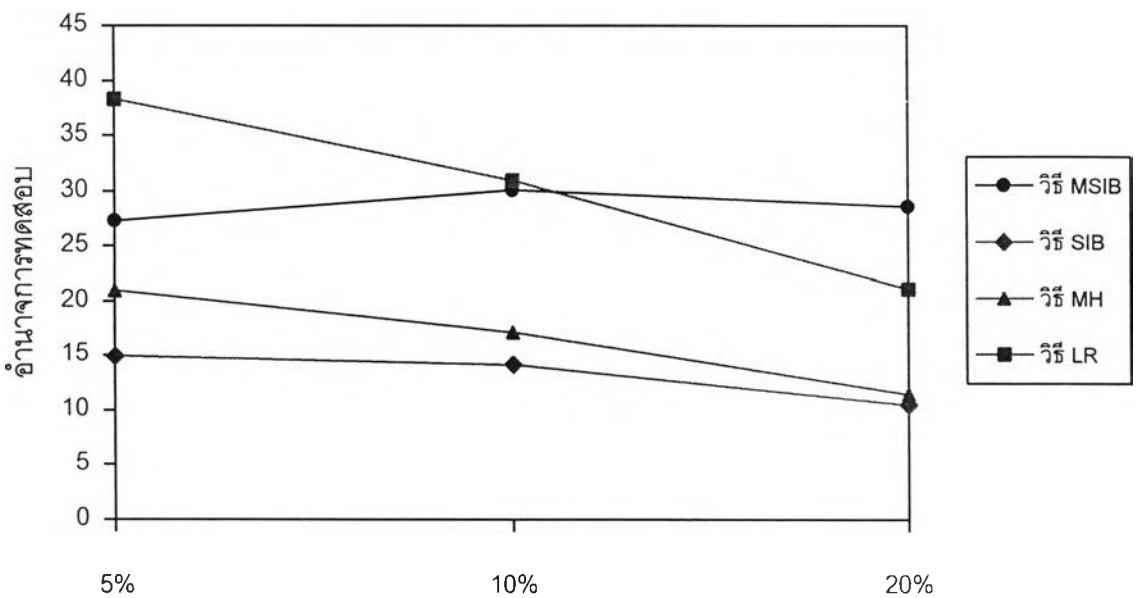
จากภาพที่ 4 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันและส่วนมากตัดกัน แสดงว่าปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบในลักษณะไม่เป็นลำดับ (disordinal interaction effect) นั่นคือทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะไม่เหมือนกันภายใต้ทุกลักษณะของข้อสอบ ดังจะเห็นได้จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_M; a_L, b_H และ a_M, b_L อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก แต่ภายใต้ลักษณะของข้อสอบ a_M, b_M; a_M, b_H; a_H, b_L และ a_H, b_M อำนาจการทดสอบของวิธีการถดถอย-โลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ เมื่อพิจารณาความลาดเอียงของกราฟ ปรากฏว่า เส้นกราฟของทั้ง 4 วิธีมีความลาดเอียงไปในทิศทางเดียวกันเกือบทุกลักษณะของข้อสอบ แสดงว่า ลักษณะของข้อสอบน่าจะมีผลต่ออำนาจการทดสอบ เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน สรุปได้ว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกลักษณะของข้อสอบ



ภาพที่ 5 กราฟแสดงการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ

จากภาพที่ 5 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันแต่มีทั้งตัดกันและไม่ตัดกัน กล่าวคือ เส้นกราฟของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกตัดกัน ส่วนเส้นกราฟของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลไม่ตัดกัน แสดงว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและความยาวของแบบสอบมีผลต่ออำนาจการทดสอบ โดยเส้นกราฟของวิธีที่ตัดกันมีผลของปฏิสัมพันธ์ในลักษณะเป็นลำดับ (disordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะไม่เหมือนกันภายใต้ทุกความยาวของแบบสอบ ดังจะเห็นได้จากภายใต้ความยาวของแบบสอบ 30 ข้อ อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ แต่ภายใต้ความยาวของแบบสอบ 60 ข้อ อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก ส่วนเส้นกราฟของวิธีที่ไม่ตัดกันมีผลของปฏิสัมพันธ์ในลักษณะเป็นลำดับ (ordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะเหมือนกันภายใต้ทุกความยาวของแบบสอบ ดังจะเห็นได้จากอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ เมื่อพิจารณาความลาดเอียงของ

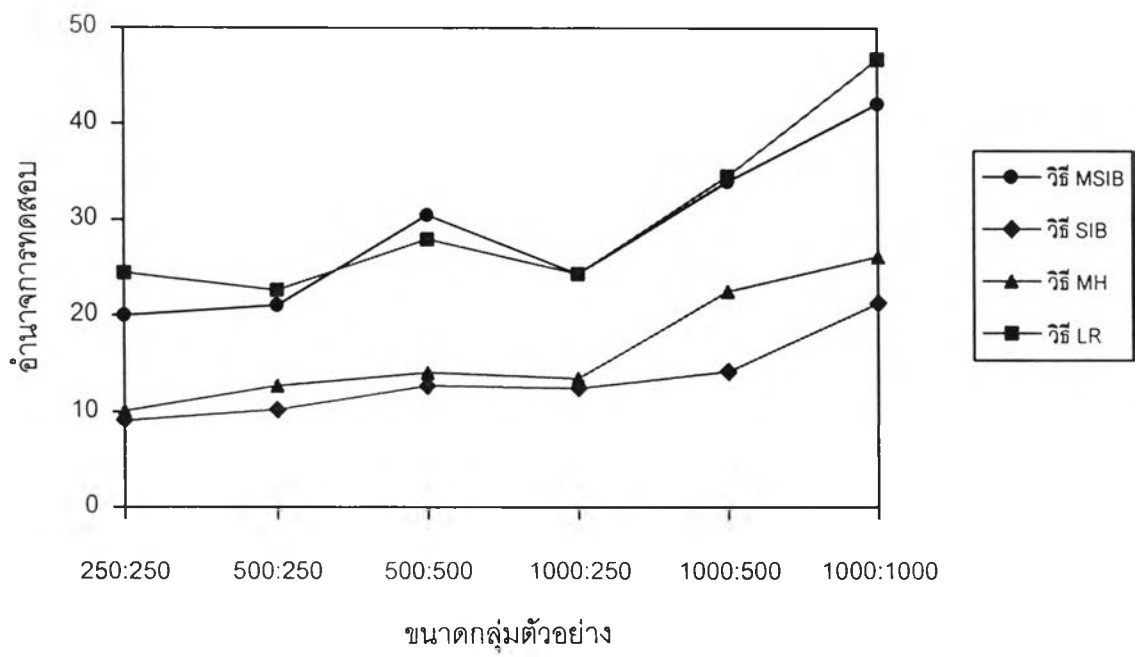
กราฟ ปรากฏว่า เส้นกราฟของวิธีการถดถอยโลจิสติกมีความลาดเอียงเพียงวิธีเดียว ส่วนของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลเกือบจะไม่มีความลาดเอียง โดยขนานกับแกนนอน แสดงว่า ความยาวของแบบสอบไม่น่าจะมีผลต่ออำนาจการทดสอบ เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน สรุปได้ว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล



ภาพที่ 6 กราฟแสดงการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

จากภาพที่ 6 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันแต่มีทั้งตัดกันและไม่ตัดกัน กล่าวคือ เส้นกราฟของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกตัดกัน ส่วนเส้นกราฟของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลไม่ตัดกัน แสดงว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีผลต่ออำนาจการทดสอบ โดยเส้นกราฟของวิธีที่ตัดกันมีผลของปฏิสัมพันธ์ในลักษณะไม่เป็นลำดับ (disordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะไม่เหมือนกันภายใต้ทุกสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ดังจะเห็นได้จากภายใต้

สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ แต่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอย-โลจิสติก ส่วนกรณีเส้นกราฟของวิธีที่ไม่ตัดกันมีผลของปฏิสัมพันธ์ในลักษณะเป็นลำดับ (ordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะเหมือนกันภายใต้ทุกสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ดังจะเห็นได้จากภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ เมื่อพิจารณาความลาดเอียงของกราฟปรากฏว่า เส้นกราฟของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลมีความ-ลาดเอียงไปในทิศทางเดียวกัน ส่วนวิธีซิปเทสท์ปรับใหม่มีความลาดเอียงไปในทิศทางตรงกันข้ามกับ 3 วิธีดังกล่าว ดังนั้นสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันน่าจะมีผลต่ออำนาจการทดสอบ เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน สรุปได้ว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล



ภาพที่ 7 กราฟแสดงการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ

จากภาพที่ 7 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันแต่มีทั้งตัดกันและไม่ตัดกัน กล่าวคือ เส้นกราฟของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกตัดกัน ส่วนเส้นกราฟของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลไม่ตัดกัน แสดงว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบโดยเส้นกราฟของวิธีที่ตัดกันมีผลของปฏิสัมพันธ์ในลักษณะไม่เป็นลำดับ (disordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะไม่เหมือนกันภายใต้ทุกขนาดกลุ่มตัวอย่าง ดังจะเห็นได้จากภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 1000 : 500 และ 1000 : 1000 อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ แต่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 500 และ 1000 : 250 อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก ส่วนกรณีเส้นกราฟของวิธีที่ไม่ตัดกันมีผลของปฏิสัมพันธ์ในลักษณะเป็นลำดับ (ordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอำนาจการทดสอบของวิธีการตรวจสอบต่างกันจะเหมือนกันภายใต้ทุกขนาดกลุ่มตัวอย่าง ดังจะเห็นได้จากอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000 เมื่อพิจารณาความลาดเอียงของกราฟ ปรากฏว่า เส้นกราฟของทั้ง 4 วิธีมีความลาดเอียงไปในทิศทางเดียวกัน ดังนั้นขนาดกลุ่มตัวอย่างน่าจะมีผลต่ออำนาจการทดสอบ เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน สรุปได้ว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล

อย่างไรก็ตาม ผลการเปรียบเทียบอำนาจการทดสอบซึ่งพิจารณาจากกราฟดังกล่าวเป็นวิธีการเปรียบเทียบอย่างคร่าว ๆ ยังไม่สามารถตอบสมมติฐานได้อย่างแท้จริง ดังนั้นผู้วิจัยจึงนำอำนาจการทดสอบของวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษา 4 ตัว จากตารางที่ 102 ในภาคผนวก ฉ หน้า 350 – 359 ไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญจะใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ทดสอบในแต่ละระดับของตัวแปร แล้วเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ยกเว้นการเปรียบเทียบภายใต้ความยาวของแบบสอบ 2 ระดับจะใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน

(independent sample) โดยจะทดสอบใน 3 กรณีต่อไปนี้ (1) ทดสอบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน (2) ทดสอบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ (3) ทดสอบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษาต่างระดับกัน ผลการวิเคราะห์ดังกล่าวแสดงดังตารางต่อไปนี้

ตารางที่ 9 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	69875.184	23291.728	48.068**
ลักษณะของข้อสอบ (TI)	8	106911.606	13363.951	27.579**
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × TI	24	45103.707	1879.321	3.878**
ความคลาดเคลื่อน	1260	610548.039	484.562	
รวม	1295	832438.536		

**p < .01

จากตารางที่ 9 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) ปรากฏว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และลักษณะของข้อสอบ (TI) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้ลักษณะของข้อสอบต่างกันจะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้นจึงนำอำนาจการทดสอบดังกล่าวไปเปรียบเทียบความแตกต่าง โดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 10

กรณีที่ 2 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 15

กรณีที่ 3 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 17

ตารางที่ 10 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี MSIB				
ระหว่างกลุ่ม	8	63627.414	7953.427	11.525**
ภายในกลุ่ม	315	217381.056	690.099	
รวม	323	281008.460		
วิธี SIB				
ระหว่างกลุ่ม	8	18501.414	2312.677	8.875**
ภายในกลุ่ม	315	82080.250	260.572	
รวม	323	100581.664		
วิธี MH				
ระหว่างกลุ่ม	8	20450.136	2556.267	6.140**
ภายในกลุ่ม	315	131149.972	416.349	
รวม	323	151600.108		

ตารางที่ 10 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี LR				
ระหว่างกลุ่ม	8	49436.351	6179.544	10.818**
ภายในกลุ่ม	315	179936.762	571.228	
รวม	323	229373.113		

**p < .01

จากตารางที่ 10 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลดังนี้

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 11

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 12

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ

อื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 13

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 14

ตารางที่ 11 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\bar{X}$	26.36	53.22	27.53	25.11	45.81	10.86	8.08	37.64	23.17
a_L, b_L	–	26.86*	1.17	1.25	19.45*	15.50	18.28	11.28	3.19
a_L, b_M		–	25.69*	28.11*	7.41	42.36*	45.14*	15.58	30.05*
a_L, b_H			–	2.42	18.28	16.67	19.45*	10.11	4.36
a_M, b_L				–	20.70*	14.25	17.03	12.53	1.94
a_M, b_M					–	34.95*	37.73*	8.17	22.64*
a_M, b_H						–	2.78	26.78*	12.31
a_H, b_L							–	29.56*	15.09
a_H, b_M								–	14.47
a_H, b_H									–

*$p < .05$

จากตารางที่ 11 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_M, b_M; a_H, b_M; a_L, b_H; a_L, b_L; a_M, b_L; a_H, b_H; a_M, b_H และ a_H, b_L ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลดังนี้

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_H; a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L และมีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M; a_L, b_H; a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H และ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ เมื่อลักษณะของข้อสอบมีค่าความยากปานกลางจะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น โดยเฉพาะการตรวจสอบภายใต้เงื่อนไข a_H, b_L อำนาจการทดสอบมีค่าประมาณร้อยละ 8.08 และภายใต้เงื่อนไข a_L, b_M อำนาจการทดสอบมีค่าประมาณร้อยละ 53.22 ดังนั้นในการตรวจสอบภายใต้ 2 เงื่อนไขดังกล่าวจะมีผลทำให้อำนาจการทดสอบมีค่าแตกต่างกันมากที่สุด ประมาณร้อยละ 45.14

ตารางที่ 12 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\overline{X}$	18.03	12.69	22.89	6.72	25.17	5.11	1.83	16.61	10.92
a_L, b_L	–	5.34	4.86	11.31	7.14	12.92*	16.20*	1.42	7.11
a_L, b_M		–	10.20	5.97	12.48*	7.58	10.86	3.92	1.77
a_L, b_H			–	16.17*	2.28	17.78*	21.06*	6.28	11.97*
a_M, b_L				–	18.45*	1.61	4.89	9.89	4.20
a_M, b_M					–	20.06*	23.34*	8.56	14.25*
a_M, b_H						–	3.28	11.50	5.81
a_H, b_L							–	14.78*	9.09
a_H, b_M								–	5.69
a_H, b_H									–

*$p < .05$

จากตารางที่ 12 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงสุด รองลงมาได้แก่ ภายใต้ลักษณะของข้อสอบ a_L, b_H; a_L, b_L; a_H, b_M; a_L, b_M; a_H, b_H; a_M, b_L; a_M, b_H และ a_H, b_L ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลดังนี้

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L และ a_M, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_H และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_H; a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกและค่าความยากปานกลางจะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยเฉพาะการตรวจสอบภายใต้เงื่อนไข a_H, b_L อำนาจการทดสอบมีค่าประมาณร้อยละ 1.83 และภายใต้เงื่อนไข a_M, b_M อำนาจการทดสอบมีค่าประมาณร้อยละ 25.17 ดังนั้นในการตรวจสอบภายใต้ 2 เงื่อนไขดังกล่าวจะมีผลทำให้อำนาจการทดสอบมีค่าแตกต่างกันมากที่สุด ประมาณร้อยละ 23.34

ตารางที่ 13 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\bar{X}$	19.17	19.22	18.53	5.78	29.14	5.53	15.97	26.08	8.61
a_L, b_L	–	0.05	0.64	13.39	9.97	13.64	3.20	6.91	10.56
a_L, b_M		–	0.69	13.44	9.92	13.69	3.25	6.86	10.61
a_L, b_H			–	12.75	10.61	13.00	2.56	7.55	9.92
a_M, b_L				–	23.36*	0.25	10.19	20.30*	2.83
a_M, b_M					–	23.61*	13.17	3.06	20.53*
a_M, b_H						–	10.44	20.55*	3.08
a_H, b_L							–	10.11	7.36
a_H, b_M								–	17.47*
a_H, b_H									–

*$p < .05$

จากตารางที่ 13 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_H, b_M; a_L, b_M; a_L, b_L; a_L, b_H; a_H, b_L; a_H, b_H; a_M, b_L และ a_M, b_H ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลดังนี้

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_M, b_H และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_M, b_H และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L และ a_M, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_M; a_L, b_H และ a_H, b_L มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ทั้งหมด

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล เมื่อลักษณะของข้อสอบมีค่าความยากปานกลางจะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น โดยเฉพาะการตรวจสอบภายใต้เงื่อนไข a_M, b_H อำนาจการทดสอบมีค่าประมาณร้อยละ 5.53 และภายใต้เงื่อนไข a_M, b_M อำนาจการทดสอบมีค่าประมาณร้อยละ 29.14 ดังนั้นในการตรวจสอบภายใต้ 2 เงื่อนไขดังกล่าวจะมีผลทำให้อำนาจการทดสอบมีค่าแตกต่างกันมากที่สุด ประมาณร้อยละ 23.61

ตารางที่ 14 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\overline{X}$	24.33	39.75	22.39	15.54	49.28	30.69	19.48	49.75	19.69
a_L, b_L	–	15.42	1.94	8.79	24.95*	6.36	4.85	25.42*	4.64
a_L, b_M		–	17.36	24.21*	9.53	9.06	20.27*	10.00	20.06*
a_L, b_H			–	6.85	26.89*	8.30	2.91	27.36*	2.70
a_M, b_L				–	33.74*	15.15	3.94	34.21*	4.15
a_M, b_M					–	18.59	29.80*	0.47	29.59*
a_M, b_H						–	11.21	19.06	11.00
a_H, b_L							–	30.27*	0.21
a_H, b_M								–	30.06*
a_H, b_H									–

*$p < .05$

จากตารางที่ 14 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_M; a_M, b_H; a_L, b_L; a_L, b_H; a_H, b_H; a_H, b_L และ a_M, b_L ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลดังนี้

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_H; a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_H; a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_H มีค่าต่ำกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_M และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก เมื่อลักษณะของข้อสอบมีค่าความยากปานกลางจะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น โดยเฉพาะการตรวจสอบภายใต้เงื่อนไข a_M, b_L และ a_H, b_M อำนาจการทดสอบมีค่าประมาณร้อยละ 15.54 และ 49.75 ตามลำดับ ดังนั้นในการตรวจสอบภายใต้ 2 เงื่อนไขดังกล่าวจะมีผลทำให้อำนาจการทดสอบมีค่าแตกต่างกันมากที่สุด ประมาณร้อยละ 34.21

ตารางที่ 15 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบเดียวกัน 9 ลักษณะ

แหล่งของความแปรปรวน	df	SS	MS	F
a_L, b_L				
ระหว่างกลุ่ม	3	1737.611	579.204	1.065
ภายในกลุ่ม	140	76166.278	544.045	
รวม	143	77903.889		
a_L, b_M				
ระหว่างกลุ่ม	3	37584.056	12528.019	22.559**
ภายในกลุ่ม	140	77748.833	555.349	
รวม	143	115332.889		
a_L, b_H				
ระหว่างกลุ่ม	3	1467.944	489.315	0.812
ภายในกลุ่ม	140	84400.056	602.858	
รวม	143	85868.000		
a_M, b_L				
ระหว่างกลุ่ม	3	8796.837	2932.279	9.733**
ภายในกลุ่ม	140	42179.830	301.284	
รวม	143	50976.677		
a_M, b_M				
ระหว่างกลุ่ม	3	15466.472	5155.491	5.426**
ภายในกลุ่ม	140	133020.167	950.144	
รวม	143	148486.639		

ตารางที่ 15 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
a_M, b_H				
ระหว่างกลุ่ม	3	15686.187	5228.729	24.105**
ภายในกลุ่ม	140	30368.472	216.918	
รวม	143	46054.659		
a_H, b_L				
ระหว่างกลุ่ม	3	6793.513	2264.504	7.5040**
ภายในกลุ่ม	140	42248.099	301.772	
รวม	143	49041.612		
a_H, b_M				
ระหว่างกลุ่ม	3	22233.576	7411.192	13.333**
ภายในกลุ่ม	140	77822.361	555.874	
รวม	143	100055.937		
a_H, b_H				
ระหว่างกลุ่ม	3	5212.694	1737.565	5.221**
ภายในกลุ่ม	140	46593.944	332.814	
รวม	143	51806.638		

**$p < .01$

จากตารางที่ 15 พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_L และ a_L, b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน ส่วนภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบดังตารางที่ 16

ตารางที่ 16 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบ a_L, b_M; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\overline{X}$	53.22	12.69	19.22	39.75
a_L, b_M	วิธี MSIB	–	40.53*	34.00*	13.47
	วิธี SIB		–	6.53	27.06*
	วิธี MH			–	20.53*
	วิธี LR				–
	$\overline{X}$	25.11	6.72	5.78	15.54
a_M, b_L	วิธี MSIB	–	18.39*	19.33*	9.57
	วิธี SIB		–	0.94	8.82
	วิธี MH			–	9.76
	วิธี LR				–
	$\overline{X}$	45.81	25.17	29.14	49.28
a_M, b_M	วิธี MSIB	–	20.64*	16.67	3.47
	วิธี SIB		–	3.97	24.11*
	วิธี MH			–	20.14*
	วิธี LR				–
	$\overline{X}$	10.86	5.11	5.53	30.69
a_M, b_H	วิธี MSIB	–	5.57	5.33	19.83*
	วิธี SIB		–	0.42	25.58*
	วิธี MH			–	20.05*
	วิธี LR				–

ตารางที่ 16 (ต่อ)

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
$\bar{X}$		8.08	1.83	15.97	19.48
a_H, b_L	วิธี MSIB	–	6.25	7.89	11.40*
	วิธี SIB		–	14.14*	17.65*
	วิธี MH			–	3.51
	วิธี LR				–
$\bar{X}$		37.64	16.61	26.08	49.75
a_H, b_M	วิธี MSIB	–	21.03*	11.56	12.11
	วิธี SIB		–	9.47	33.14*
	วิธี MH			–	23.67*
	วิธี LR				–
$\bar{X}$		23.17	10.92	8.61	19.69
a_H, b_H	วิธี MSIB	–	12.25*	14.56*	3.48
	วิธี SIB		–	2.31	8.77
	วิธี MH			–	11.08*
	วิธี LR				–

*p < .05

จากตารางที่ 16 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยลักษณะของข้อสอบ a_L, b_M; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H ปรากฏผลดังนี้

ภายใต้ลักษณะของข้อสอบ a_L, b_M อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอย-โลจิสติก วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่า

อำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_L อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_M อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์-ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_L อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีแมนเทล-แฮนส์เซล วิธีซิปเทสท์ปรับใหม่ และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าสูงกว่าอำนาจการทดสอบของ

วิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_M อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์-ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอย-โลจิสติก วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอย-โลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไขของลักษณะของข้อสอบ

ตารางที่ 17 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ

วิธีการตรวจสอบ DIF (TI) $\bar{X}$	MSIB (a_L, b_L) 26.36	MSIB (a_L, b_M) 53.22	MSIB (a_L, b_H) 27.53	MSIB (a_M, b_L) 25.11	MSIB (a_M, b_M) 45.81	MSIB (a_M, b_H) 10.86	MSIB (a_H, b_L) 8.08	MSIB (a_H, b_M) 37.64	MSIB (a_H, b_H) 23.17	SIB (a_L, b_L) 18.03	SIB (a_L, b_M) 12.69	SIB (a_L, b_H) 22.89	SIB (a_M, b_L) 6.72	SIB (a_M, b_M) 25.17	SIB (a_M, b_H) 5.11	SIB (a_H, b_L) 1.83	SIB (a_H, b_M) 16.61	SIB (a_H, b_H) 10.92
MSIB (a_L, b_L)	–	26.86*	1.17	1.25	19.45	15.50	18.28	11.28	3.19	8.33	13.67	3.47	19.64	1.19	21.25*	24.53*	9.75	15.44
MSIB (a_L, b_M)		–	25.69*	28.11*	7.41	42.36*	45.14*	15.58	30.05*	35.19*	40.53*	30.33*	46.50*	28.05*	48.11*	51.39*	36.61*	42.30*
MSIB (a_L, b_H)			–	2.42	18.28	16.67	19.45	10.11	4.36	9.50	14.84	4.64	20.81*	2.36	22.42*	25.70*	10.92	16.61
MSIB (a_M, b_L)				–	20.70*	14.25	17.03	12.53	1.94	7.08	12.42	2.22	18.39	0.06	20.00*	23.28*	8.50	14.19
MSIB (a_M, b_M)					–	34.95*	37.73*	8.17	22.6*4	27.78*	33.12*	22.92*	39.09*	20.64*	40.70*	43.98*	29.20*	34.89*
MSIB (a_M, b_H)						–	2.78	26.78*	12.31	7.17	1.83	12.03	4.14	14.31	5.75	9.03	5.75	0.06
MSIB (a_H, b_L)							–	29.56*	15.09	9.95	4.61	14.81	1.36	17.09	2.97	6.25	8.53	2.84
MSIB (a_H, b_M)								–	14.47	19.61	24.95*	14.75	30.92*	12.47	32.53*	35.81*	21.03*	26.72*
MSIB (a_H, b_H)									–	5.14	10.48	0.28	16.45	2.00	18.06	21.34*	6.56	12.25
SIB (a_L, b_L)										–	5.34	4.86	11.31	7.14	12.92	16.20	1.42	7.11
SIB (a_L, b_M)											–	10.20	5.97	12.48	7.58	10.86	3.92	1.77
SIB (a_L, b_H)												–	16.17	2.28	17.78	21.06*	6.28	11.97
SIB (a_M, b_L)													–	18.45	1.61	4.89	9.89	4.20
SIB (a_M, b_M)														–	20.06*	23.34*	8.56	14.25
SIB (a_M, b_H)															–	3.28	11.50	5.81
SIB (a_H, b_L)																–	14.78	9.09
SIB (a_H, b_M)																	–	5.69
SIB (a_H, b_H)																		–

ตารางที่ 17 (ต่อ)

วิธีการตรวจสอบ DIF (TI) $\bar{X}$	MH (a_L, b_L) 19.17	MH (a_L, b_M) 19.22	MH (a_L, b_H) 18.53	MH (a_M, b_L) 5.78	MH (a_M, b_M) 29.14	MH (a_M, b_H) 5.53	MH (a_H, b_L) 15.97	MH (a_H, b_M) 26.08	MH (a_H, b_H) 8.61	LR (a_L, b_L) 24.33	LR (a_L, b_M) 39.75	LR (a_L, b_H) 22.39	LR (a_M, b_L) 15.54	LR (a_M, b_M) 49.28	LR (a_M, b_H) 30.69	LR (a_H, b_L) 19.48	LR (a_H, b_M) 49.75	LR (a_H, b_H) 19.69
MSIB (a_L, b_L)	7.19	7.14	7.83	20.58*	2.78	20.83*	10.39	0.28	17.75	2.03	13.39	3.97	10.82	22.92*	4.33	6.88	23.39*	6.67
MSIB (a_L, b_M)	34.05*	34.00*	34.69*	47.44*	24.08*	47.69*	37.25*	27.14*	44.61*	28.89*	13.47	30.83*	37.68*	3.94	22.53*	33.74*	3.47	33.53*
MSIB (a_L, b_H)	8.36	8.31	9.00	21.75*	1.61	22.00*	11.56	1.45	18.92	3.20	12.22	5.14	11.99	21.75*	3.16	8.05	22.22*	7.84
MSIB (a_M, b_L)	5.94	5.89	6.58	19.33	4.03	19.58	9.14	0.97	16.50	0.78	14.64	2.72	9.57	24.17*	5.58	5.63	24.64*	5.42
MSIB (a_M, b_M)	26.64*	26.59*	27.28*	40.03*	16.67	40.28*	29.84*	19.73	37.20*	21.48*	6.06	23.42*	30.27*	3.47	15.12	26.33*	3.94	26.12*
MSIB (a_M, b_H)	8.31	8.36	7.67	5.08	18.28	5.33	5.11	15.22	2.25	13.47	28.89*	11.53	4.68	38.42*	19.83	8.62	38.89*	8.83
MSIB (a_H, b_L)	11.09	11.14	10.45	2.30	21.06*	2.55	7.89	18.00	0.53	16.25	31.67*	14.31	7.46	41.20*	22.61*	11.40	41.67*	11.61
MSIB (a_H, b_M)	18.47	18.42	19.11	31.86*	8.50	32.11*	21.67*	11.56	29.03*	13.31	2.11	15.25	22.10*	11.64	6.95	18.16	12.11	17.95
MSIB (a_H, b_H)	4.00	3.95	4.64	17.39	5.97	17.64	7.20	2.91	14.56	1.16	16.58	0.78	7.63	26.11*	7.52	3.69	26.58*	3.48
SIB (a_L, b_L)	1.14	1.19	0.50	12.25	11.11	12.50	2.06	8.05	9.42	6.30	21.72*	4.36	2.49	31.25*	12.66	1.45	31.72*	1.66
SIB (a_L, b_M)	6.48	6.53	5.84	6.91	16.45	7.16	3.28	13.39	4.08	11.64	27.06*	9.70	2.85	36.59*	18.00	6.79	37.06*	7.00
SIB (a_L, b_H)	3.72	3.67	4.36	17.11	6.25	17.36	6.92	3.19	14.28	1.44	16.86	0.50	7.35	26.39*	7.80	3.41	26.86*	3.20
SIB (a_M, b_L)	12.45	12.50	11.81	0.94	22.42*	1.19	9.25	19.36	1.89	17.61	33.03*	15.67	8.82	42.56*	23.97*	12.76	43.03*	12.97
SIB (a_M, b_M)	6.00	5.95	6.64	19.39	3.97	19.64	9.20	0.91	16.56	0.84	14.58	2.78	9.63	24.11*	5.52	5.69	24.58*	5.48
SIB (a_M, b_H)	14.06	14.11	13.42	0.67	24.03*	0.42	10.86	20.97*	3.50	19.22	34.64*	17.28	10.43	44.17*	25.58*	14.37	44.64*	14.58
SIB (a_H, b_L)	17.34	17.39	16.70	3.95	27.31*	3.70	14.14	24.25*	6.78	22.50*	37.92*	20.56*	13.71	47.45*	28.86*	17.65	47.92*	17.86
SIB (a_H, b_M)	2.56	2.61	1.92	10.83	12.53	11.08	0.64	9.47	8.00	7.72	23.14*	5.78	1.07	32.67*	24.08*	2.87	33.14*	3.08
SIB (a_H, b_H)	8.25	8.30	7.61	5.14	18.22	5.36	5.05	15.16	2.31	13.41	28.83*	11.47	4.62	38.36*	19.77	0.29	38.83*	8.77

ตารางที่ 17 (ต่อ)

วิธีการตรวจสอบ DIF (TI)	MH (a_L, b_L)	MH (a_L, b_M)	MH (a_L, b_H)	MH (a_M, b_L)	MH (a_M, b_M)	MH (a_M, b_H)	MH (a_H, b_L)	MH (a_H, b_M)	MH (a_H, b_H)	LR (a_L, b_L)	LR (a_L, b_M)	LR (a_L, b_H)	LR (a_M, b_L)	LR (a_M, b_M)	LR (a_M, b_H)	LR (a_H, b_L)	LR (a_H, b_M)	LR (a_H, b_H)
$\bar{X}$	19.17	19.22	18.53	5.78	29.14	5.53	15.97	26.08	8.61	24.33	39.75	22.39	15.54	49.28	30.69	19.48	49.75	19.69
MH (a_L, b_L)	–	0.05	0.64	13.39	9.97	13.64	3.20	6.91	10.56	5.16	20.58*	3.22	3.63	30.11*	11.52	0.31	30.58*	0.52
MH (a_L, b_M)		–	0.69	13.44	9.92	13.69	3.25	6.86	10.61	5.11	20.53*	3.17	3.68	30.06*	11.47	0.26	30.53*	0.47
MH (a_L, b_H)			–	12.75	10.61	13.00	2.56	7.55	9.92	5.80	21.22*	3.86	2.99	30.75*	12.16	0.95	31.22*	1.16
MH (a_M, b_L)				–	23.36*	0.25	10.19	20.30*	2.83	18.55	33.97*	16.61	9.76	43.50*	24.91*	13.70	43.97*	13.91
MH (a_M, b_M)					–	23.61*	13.17	3.06	20.53*	4.81	10.61	6.75	13.60	20.14*	1.55	9.66	20.61*	9.45
MH (a_M, b_H)						–	10.44	20.55*	3.08	18.80	34.22*	16.86	10.01	43.75*	25.16*	13.95	44.22*	14.16
MH (a_H, b_L)							–	10.11	7.36	8.36	23.78*	6.42	0.43	33.31*	14.72	3.51	33.78*	3.72
MH (a_H, b_M)								–	17.47	1.75	13.67	3.69	10.54	23.20*	4.61	6.60	23.67*	6.39
MH (a_H, b_H)									–	15.72	31.14*	13.78	6.93	40.67*	22.08*	10.87	41.14*	11.08
LR (a_L, b_L)										–	15.42	1.94	8.79	24.95*	6.36	17.97	25.42*	4.64
LR (a_L, b_M)											–	17.36	24.21*	9.53	9.06	20.37*	10.00	20.06*
LR (a_L, b_H)												–	6.85	26.89*	8.30	2.91	27.36*	2.70
LR (a_M, b_L)													–	33.74*	15.15	3.94	34.21*	4.15
LR (a_M, b_M)														–	18.59	29.80*	0.47	29.59*
LR (a_M, b_H)															–	11.21	19.06	11.00
LR (a_H, b_L)																–	30.27*	0.21
LR (a_H, b_M)																	–	30.06*
LR (a_H, b_H)																		–

*$p < .05$

จากตารางที่ 17 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้ปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยเปรียบเทียบค่าเฉลี่ยของอำนาจการทดสอบซึ่งมีทั้งหมด 36 ค่า เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีMSIB(a_L,b_M)มีค่าสูงสุด รองลงมาได้แก่ วิธีLR(a_H,b_M)และวิธีLR(a_M,b_M) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีSIB(a_H,b_L)มีค่าต่ำสุด รองลงมาได้แก่ วิธีSIB(a_M,b_H)และวิธีMH (a_M,b_H) ตามลำดับ ผลการเปรียบเทียบอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบดังกล่าวกับอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ปรากฏผล ดังนี้

อำนาจการทดสอบของวิธีMSIB(a_L,b_M)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีMSIB(a_M,b_M), วิธีMSIB(a_H,b_M), วิธีLR(a_L,b_M), วิธีLR(a_M,b_M) และวิธีLR(a_H,b_M) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีLR(a_H,b_M)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีMSIB(a_L,b_M), วิธีMSIB(a_M,b_M), วิธีMSIB(a_H,b_M), วิธีLR(a_L,b_M), วิธีLR(a_M,b_M) และวิธีLR (a_M,b_H) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีLR(a_M,b_M)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีMSIB(a_L,b_M), วิธีMSIB(a_M,b_M), วิธีMSIB(a_H,b_M), วิธีLR(a_L,b_M), วิธีLR(a_M,b_H), และวิธีLR (a_H,b_M) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีSIB(a_H,b_L)มีค่าต่ำกว่าอำนาจการทดสอบของวิธีMSIB (a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB(a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB (a_H,b_M), วิธีMSIB(a_H,b_H), วิธีSIB(a_L,b_H), วิธีSIB(a_M,b_M), วิธีMH(a_M,b_M), วิธีMH(a_H,b_M), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_M), วิธีLR(a_M,b_H) และวิธีLR(a_H,b_M) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีSIB(a_M,b_H)มีค่าต่ำกว่าอำนาจการทดสอบของวิธีMSIB (a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB(a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB (a_H,b_M), วิธีSIB(a_M,b_M), วิธีMH(a_M,b_M), วิธีMH(a_H,b_M), วิธีLR(a_L,b_M), วิธีLR(a_M,b_M), วิธีLR(a_M,b_H) และวิธีLR(a_H,b_M) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีMH(a_M,b_H)มีค่าต่ำกว่าอำนาจการทดสอบของวิธีMSIB (a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB(a_M,b_M), วิธีMSIB(a_H,b_M), วิธีMH(a_M,b_M), วิธีMH(a_H,b_M), วิธีLR(a_L,b_M), วิธีLR(a_M,b_M), วิธีLR(a_M,b_H) และวิธีLR(a_H,b_M) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป เมื่อใช้วิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a ต่ำกับ b ปานกลาง จะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ในทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบ a สูงกับ b ต่ำจะส่งผลให้อำนาจการทดสอบมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ

ตารางที่ 18 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	69875.184	23291.728	39.375**
ความยาวของแบบสอบ (TL)	1	57.942	57.942	.098
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × TL	3	603.919	201.306	.340
ความคลาดเคลื่อน	1288	761901.493	591.538	
รวม	1295	832438.538		

**p < .01

จากตารางที่ 18 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) ปรากฏว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยาวของแบบสอบ (TL) ไม่มีนัยสำคัญทางสถิติ แสดงว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้ความยาวของแบบสอบต่างกันแล้วอำนาจการทดสอบมีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของปัจจัยความยาวของแบบสอบ (TL) พบว่า ความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ความยาวของแบบสอบต่างกันแล้วอำนาจการทดสอบจะมีค่าไม่แตกต่างกัน นั่นคือ ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบ DIF มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันจะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้นจึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน 2

ระดับไปเปรียบเทียบความแตกต่าง โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 19

ตารางที่ 19 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบระดับเดียวกัน 2 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
30 ข้อ				
ระหว่างกลุ่ม	3	37740.907	12580.302	18.738**
ภายในกลุ่ม	644	432372.085	671.385	
รวม	647	470112.992		
60 ข้อ				
ระหว่างกลุ่ม	3	32738.196	10912.732	21.327**
ภายในกลุ่ม	644	329529.408	511.692	
รวม	647	362267.604		

**p < .01

จากตารางที่ 19 ปรากฏว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้ความยาวของแบบสอบทั้ง 2 ระดับไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 20

ตารางที่ 20 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบระดับเดียวกัน 2 ระดับ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\bar{X}$	28.17	13.05	16.77	31.38
30 ข้อ	วิธี MSIB	–	15.12*	11.40*	3.21
	วิธี SIB		–	3.72	18.33*
	วิธี MH			–	14.61*
	วิธี LR				–
	$\bar{X}$	29.12	13.61	16.12	28.82
60 ข้อ	วิธี MSIB	–	15.51*	13.00*	0.30
	วิธี SIB		–	2.51	15.21*
	วิธี MH			–	12.70*
	วิธี LR				–

*p < .05

จากตารางที่ 20 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ ปรากฏผลดังนี้

ภายใต้ความยาวของแบบสอบ 30 ข้อ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์-ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการ-ถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ความยาวของแบบสอบ 60 ข้อ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้ทุกเงื่อนไขของความยาวของแบบสอบ

ตารางที่ 21 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอำนาจการทดสอบ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	69875.184	23291.728	40.418**
สัดส่วนของ DIF (%DIF)	2	12492.125	6246.063	10.839**
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × %DIF	6	10143.088	1690.515	2.934**
ความคลาดเคลื่อน	1284	739928.141	576.268	
รวม	1295	832438.538		

**p < .01

จากตารางที่ 21 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้นจึงเปรียบเทียบอำนาจการทดสอบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 22

กรณีที่ 2 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 25

กรณีที่ 3 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 27

ตารางที่ 22 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี MSIB				
ระหว่างกลุ่ม	2	423.451	211.725	.242
ภายในกลุ่ม	321	280585.019	874.097	
รวม	323	281008.469		

ตารางที่ 22 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี SIB				
ระหว่างกลุ่ม	2	1025.451	512.725	1.653
ภายในกลุ่ม	321	99556.213	310.144	
รวม	323			
วิธี MH				
ระหว่างกลุ่ม	2	5111.710	2555.855	5.601**
ภายในกลุ่ม	321	146488.398	456.350	
รวม	323	151600.108		
วิธี LR				
ระหว่างกลุ่ม	2	16074.602	8037.301	12.096**
ภายในกลุ่ม	321	213298.511	664.481	
รวม	323	229373.113		

**p < .01

จากตารางที่ 22 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ ปรากฏผลดังนี้

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับมีค่าไม่แตกต่างกัน แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับมีค่าไม่แตกต่างกัน แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบดังกล่าวไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 23

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีการถดถอย-โลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบดังกล่าวไปเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 24

ตารางที่ 23 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Tukey

%DIF	5%	10%	20%
$\bar{X}$	20.94	17.13	11.28
5%	–	3.81	9.66*
10%		–	5.85
20%			–

*p < .05

จากตารางที่ 23 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงสุด รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ตามลำดับ

เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงกว่าอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันจาก 5% เป็น 20% จะมีผลทำให้อำนาจการทดสอบมีค่าลดลงต่ำสุด ประมาณร้อยละ 9.66

ตารางที่ 24 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Tukey

%DIF	5%	10%	20%
$\bar{X}$	38.24	31.00	21.06
5%	–	7.24	17.18*
10%		–	9.94*
20%			–

*p < .05

จากตารางที่ 24 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงสุด รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ ปรากฏว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% มีค่าไม่แตกต่างกัน แต่ทั้ง 2 ระดับมีอำนาจการทดสอบสูงกว่าภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันจาก 5% เป็น 20% จะมีผลทำให้อำนาจการทดสอบมีค่าลดลงต่ำสุด ประมาณร้อยละ 13.63

ตารางที่ 25 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
5%				
ระหว่างกลุ่ม	3	32150.463	10716.821	12.056**
ภายในกลุ่ม	428	380452.056	888.907	
รวม	431	412602.519		
10%				
ระหว่างกลุ่ม	3	24427.380	8142.460	14.071**
ภายในกลุ่ม	428	247671.000	578.671	
รวม	431	272098.380		
20%				
ระหว่างกลุ่ม	3	23440.428	7813.476	29.911**
ภายในกลุ่ม	428	111805.085	261.227	
รวม	431	135245.514		

**p < .01

จากตารางที่ 25 พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกัน

อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 ระดับดังกล่าวไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 26

ตารางที่ 26 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\bar{X}$	27.29	14.94	20.94	38.24
5%	วิธี MSIB	–	12.34*	6.35	10.95*
	วิธี SIB		–	6.00	23.30*
	วิธี MH			–	17.30*
	วิธี LR				–
	$\bar{X}$	30.08	14.19	17.13	31.00
10%	วิธี MSIB	–	15.89*	12.95*	.92
	วิธี SIB		–	2.94	16.81*
	วิธี MH			–	13.87*
	วิธี LR				–
	$\bar{X}$	28.56	10.85	11.28	21.06
20%	วิธี MSIB	–	17.71*	17.28*	7.50*
	วิธี SIB		–	0.43	10.21*
	วิธี MH			–	9.78*
	วิธี LR				–

*p < .05

จากตารางที่ 26 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ ปรากฏผลดังนี้

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของอีก 3 วิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีซิปเทสท์-ปรับใหม่มีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของอีก 3 วิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันจำนวน 5%, 10% และ 20% อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 27 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

วิธีการตรวจสอบ	MSIB	MSIB	MSIB	SIB	SIB	SIB	MH	MH	MH	LR	LR	LR
DIF (TI)	(5%)	(10%)	(20%)	(5%)	(10%)	(20%)	(5%)	(10%)	(20%)	(5%)	(10%)	(20%)
$\bar{X}$	27.29	30.08	28.56	14.94	14.19	10.85	20.94	17.13	11.28	38.24	31.00	21.06
MSIB (5%)	–	2.79	1.27	12.35*	13.10*	16.44*	6.35	10.16	16.01*	10.95*	3.71	6.23
MSIB (10%)		–	1.52	15.14*	15.89*	19.23*	9.14	12.95*	18.80*	8.16	0.92	9.02
MSIB (20%)			–	13.62*	14.37*	17.71*	7.62	11.43*	17.28*	9.68	2.44	7.50
SIB (5%)				–	0.75	4.09	6.00	2.19	3.66	23.30*	16.06*	6.12
SIB (10%)					–	3.43	6.75	2.94	2.91	24.05*	16.81*	6.87
SIB (20%)						–	10.09	6.28	0.43	27.39*	20.15*	10.21
MH (5%)							–	3.81	9.66	17.30*	10.06	0.12
MH (10%)								–	5.85	21.11*	13.87*	3.93
MH (20%)									–	26.96*	19.72*	9.78
LR (5%)										–	7.24	17.18*
LR (10%)											–	9.94
LR (20%)												–

*p < .05

จากตารางที่ 27 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ โดยเปรียบเทียบค่าเฉลี่ยของอำนาจการทดสอบซึ่งมีทั้งหมด 12 ค่า เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีLR(5%)มีค่าสูงสุด รองลงมาได้แก่ วิธีLR(10%)และวิธีMSIB(10%) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีSIB(20%)มีค่าต่ำสุด รองลงมาได้แก่ วิธีMH(20%)และวิธีSIB(10%) ตามลำดับ ผลการเปรียบเทียบอำนาจการทดสอบของวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันดังกล่าวกับวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ปรากฏผล ดังนี้

อำนาจการทดสอบของวิธีLR(5%)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีMSIB (10%), วิธีMSIB(20%) และวิธีLR(10%) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีLR(10%)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีMSIB (5%), วิธีMSIB(10%), วิธีMSIB(20%), วิธีMH(5%) และวิธีLR(5%) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีMSIB(10%)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธี MSIB(5%), วิธีMSIB(20%), วิธีMH(5%), วิธีLR(5%), วิธีLR(10%) และวิธีLR(20%) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีSIB(20%)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีSIB (5%), วิธีSIB(10%), วิธีMH(5%), วิธีMH(10%), วิธีMH(20%) และวิธีLR(20%) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีMH(20%)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีSIB (5%), วิธีSIB(10%), วิธีSIB(20%), วิธีMH(5%), วิธีMH(10%) และวิธีLR(20%) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อำนาจการทดสอบของวิธีSIB(10%)มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีSIB (5%), วิธีSIB(20%), วิธีMH(5%), วิธีMH(10%), วิธีMH(20%) และวิธีLR(20%) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป เมื่อใช้วิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% จะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ ในขณะเดียวกัน เมื่อใช้วิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะส่งผลให้อำนาจการทดสอบมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ

ตารางที่ 28 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	69875.184	23291.728	42.060**
ขนาดกลุ่มตัวอย่าง (N)	5	51834.544	10366.909	18.721**
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × N	15	6333.347	422.223	.762
ความคลาดเคลื่อน	1272	704395.463	553.770	
รวม	1295	832438.538		

**p < .01

จากตารางที่ 28 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) ปรากฏว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และขนาดกลุ่มตัวอย่าง (N) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้ขนาดกลุ่มตัวอย่างต่างกันแล้วอำนาจการทดสอบมีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและขนาดกลุ่มตัวอย่างไม่มีผลร่วมกันต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของปัจจัยขนาดกลุ่มตัวอย่าง (N) พบว่า ขนาดกลุ่มตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ขนาดกลุ่มตัวอย่างต่างกันจะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้นจึงนำอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 29

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบ DIF มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันจะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้นจึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 34

ตารางที่ 29 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี MSIB				
ระหว่างกลุ่ม	5	19649.136	3929.827	4.781**
ภายในกลุ่ม	318	261359.333	821.885	
รวม	323	281008.469		

ตารางที่ 29 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี SIB				
ระหว่างกลุ่ม	5	5010.386	1002.077	4.781**
ภายในกลุ่ม	318	95571.278	821.885	
รวม	323	100581.664		
วิธี MH				
ระหว่างกลุ่ม	8	5010.386	1002.077	3.334**
ภายในกลุ่ม	315	95571.278	300.539	
รวม	323	100581.664		
วิธี LR				
ระหว่างกลุ่ม	5	22862.946	4572.589	7.041**
ภายในกลุ่ม	318	206510.166	649.403	
รวม	323	229373.113		

**p < .01

จากตารางที่ 29 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ ปรากฏผลดังนี้

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างในระดับอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 30

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างในระดับอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 31

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างในระดับอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 32

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างในระดับอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 33

ตารางที่ 30 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่ม-ตัวอย่างต่างกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

$N_R : N_F$	250 : 250	500 : 250	500 : 500	1000 : 250	1000 : 500	1000 : 1000
$\bar{X}$	20.00	21.09	30.41	24.28	33.94	42.13
250 : 250	–	1.09	10.41	4.28	13.94	22.13*
500 : 250		–	9.32	3.19	12.85	21.04*
500 : 500			–	6.13	3.53	11.72
1000 : 250				–	9.66	17.85*
1000 : 500					–	8.19
1000 :1000						–

*p < .05

จากตารางที่ 30 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 500 : 500, 1000 : 250, 500 : 250 และ 250 : 250 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250 และ 1000 : 250 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก $N_R : N_F$ = 250 : 250 เป็น $N_R : N_F$ = 1000 : 1000 จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้นสูงสุด ประมาณร้อยละ 22.13

ตารางที่ 31 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

$N_R : N_F$	250 : 250	500 : 250	500 : 500	1000 : 250	1000 : 500	1000 : 1000
$\bar{X}$	9.09	10.17	12.63	12.65	14.17	21.28
250 : 250	–	1.08	3.54	3.56	5.08	12.19*
500 : 250		–	2.46	2.48	4.00	11.11*
500 : 500			–	0.02	1.54	8.65
1000 : 250				–	1.52	8.63
1000 : 500					–	7.11
1000 :1000						–

*p < .05

จากตารางที่ 31 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 1000 : 250, 500 : 500, 500 : 250 และ 250 : 250 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 และ 500 : 250 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก $N_R : N_F$ = 250 : 250 เป็น $N_R : N_F$ = 1000 : 1000 จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้นสูงสุด ประมาณร้อยละ 12.19

ตารางที่ 32 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่ม-ตัวอย่างต่างกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

$N_R : N_F$	250 : 250	500 : 250	500 : 500	1000 : 250	1000 : 500	1000 : 1000
$\bar{X}$	10.80	12.72	13.09	13.41	22.54	26.13
250 : 250	–	1.92	2.29	2.61	11.74	15.33*
500 : 250		–	0.37	0.69	9.82	13.41*
500 : 500			–	0.32	9.45	13.04*
1000 : 250				–	9.13	12.72*
1000 : 500					–	3.59
1000 :1000						–

*$p < .05$

จากตารางที่ 32 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 1000 : 250, 500 : 500, 500 : 250 และ 250 : 250 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500 และ 1000 : 250 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500 และ 1000 : 250 มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก $N_R : N_F$ = 250 : 250 เป็น $N_R : N_F$ = 1000 : 1000 จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้นสูงสุด ประมาณร้อยละ 15.33

ตารางที่ 33 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

$N_R : N_F$	250 : 250	500 : 250	500 : 500	1000 : 250	1000 : 500	1000 : 1000
$\bar{X}$	24.38	22.63	27.93	24.37	34.54	46.76
250 : 250	–	1.75	3.55	0.01	10.16	22.38*
500 : 250		–	5.30	1.74	11.91	24.13*
500 : 500			–	3.56	6.61	18.83*
1000 : 250				–	10.17	22.39*
1000 : 500					–	12.22
1000 :1000						–

*p < .05

จากตารางที่ 33 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 500 : 500, 250 : 250, 1000 : 250 และ 500 : 250 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500 และ 1000 : 250 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก $N_R : N_F$ = 500 : 250 เป็น $N_R : N_F$ = 1000 : 1000 จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้นสูงสุด ประมาณร้อยละ 24.13

ตารางที่ 34 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่ม-ตัวอย่างระดับเดียวกัน 6 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
$N_R : N_F = 250 : 250$				
ระหว่างกลุ่ม	3	8690.687	2896.896	9.098**
ภายในกลุ่ม	212	67505.911	318.424	
รวม	215	76196.598		
$N_R : N_F = 500 : 250$				
ระหว่างกลุ่ม	3	6099.495	2033.165	5.286**
ภายในกลุ่ม	212	81547.463	384.658	
รวม	215	87646.958		
$N_R : N_F = 500 : 500$				
ระหว่างกลุ่ม	3	14529.088	4843.029	9.118**
ภายในกลุ่ม	212	112601.870	531.141	
รวม	215	127130.958		
$N_R : N_F = 1000 : 250$				
ระหว่างกลุ่ม	3	6906.537	2302.179	4.996**
ภายในกลุ่ม	212	97688.778	460.796	
รวม	215	104595.315		
$N_R : N_F = 1000 : 500$				
ระหว่างกลุ่ม	3	15538.798	5179.599	6.361**
ภายในกลุ่ม	212	172620.551	814.248	
รวม	215	188159.349		

ตารางที่ 34 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
$N_R : N_F$ = 1000 : 1000				
ระหว่างกลุ่ม	3	24443.926	8147.975	10.018**
ภายในกลุ่ม	212	172430.889	813.353	
รวม	215	196874.815		

**p < .01

จากตารางที่ 34 พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 :500 และ 1000 : 1000 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้ขนาดกลุ่มตัวอย่างทั้ง 6 ระดับดังกล่าวไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 35

ตารางที่ 35 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่ม-ตัวอย่างระดับเดียวกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\bar{X}$	20.00	9.09	10.80	24.38
250 : 250	วิธี MSIB	–	10.91*	9.20*	4.38
	วิธี SIB		–	1.71	15.29*
	วิธี MH			–	13.58*
	วิธี LR				–

ตารางที่ 35 (ต่อ)

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\overline{X}$	21.09	10.17	12.72	22.63
500 : 250	วิธี MSIB	–	10.92*	8.37	1.54
	วิธี SIB		–	2.55	12.46*
	วิธี MH			–	9.91*
	วิธี LR				–
	$\overline{X}$	30.41	12.63	13.09	27.93
500 : 500	วิธี MSIB	–	17.78*	17.32*	2.48
	วิธี SIB		–	0.46	15.30*
	วิธี MH			–	14.84*
	วิธี LR				–
	$\overline{X}$	24.28	12.65	13.41	24.37
1000 : 250	วิธี MSIB	–	11.63*	10.87*	.09
	วิธี SIB		–	0.76	11.72*
	วิธี MH			–	10.96*
	วิธี LR				–
	$\overline{X}$	33.94	14.17	22.54	34.54
1000 : 500	วิธี MSIB	–	19.77*	11.40	.60
	วิธี SIB		–	8.37	20.37
	วิธี MH			–	12.00
	วิธี LR				–

ตารางที่ 35 (ต่อ)

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
$\overline{X}$		42.13	21.28	26.31	46.76
1000 : 1000	วิธี MSIB	–	20.85*	16.00	4.63
	วิธี SIB		–	4.85	25.48*
	วิธี MH			–	20.63*
	วิธี LR				–

*p < .05

จากตารางที่ 35 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ ปรากฏผลดังนี้

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 1000 : 250 และ 1000 : 1000 ได้ผลสอดคล้องกัน กล่าวคือ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 250 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 500 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

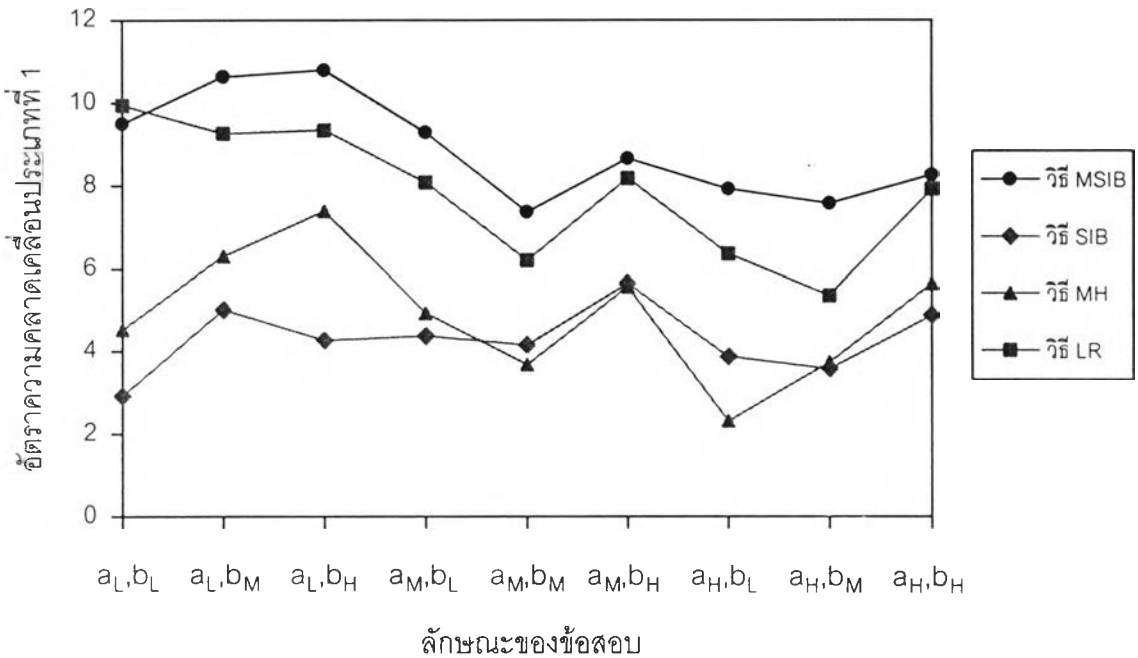
ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000 อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

2. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา

ก่อนที่จะนำผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 จากตารางที่ 8 ไปทดสอบทางสถิติ ผู้วิจัยนำผลการวิเคราะห์ดังกล่าวไปสร้างกราฟ เพื่อพิจารณาผลกระทบอย่างคร่าว ๆ ของตัวแปรอิสระที่มีต่อตัวแปรตาม ตัวแปรอิสระมี 2 ตัวหลัก คือ (1) วิธีการตรวจสอบการ-ทำหน้าที่ต่างกันของข้อสอบ 4 วิธี ได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล วิธีการถดถอยโลจิสติก และ (2) ปัจจัยที่ศึกษา 4 ตัว ได้แก่ ลักษณะของข้อสอบ 9 ลักษณะ (a_L, b_L; a_L, b_M; a_L, b_H; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H) ความยาวของแบบสอบ 2 ระดับ (30 ข้อ และ 60 ข้อ) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ (5%, 10% และ 20%) และขนาดกลุ่มตัวอย่าง 6 ระดับ ($N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500,

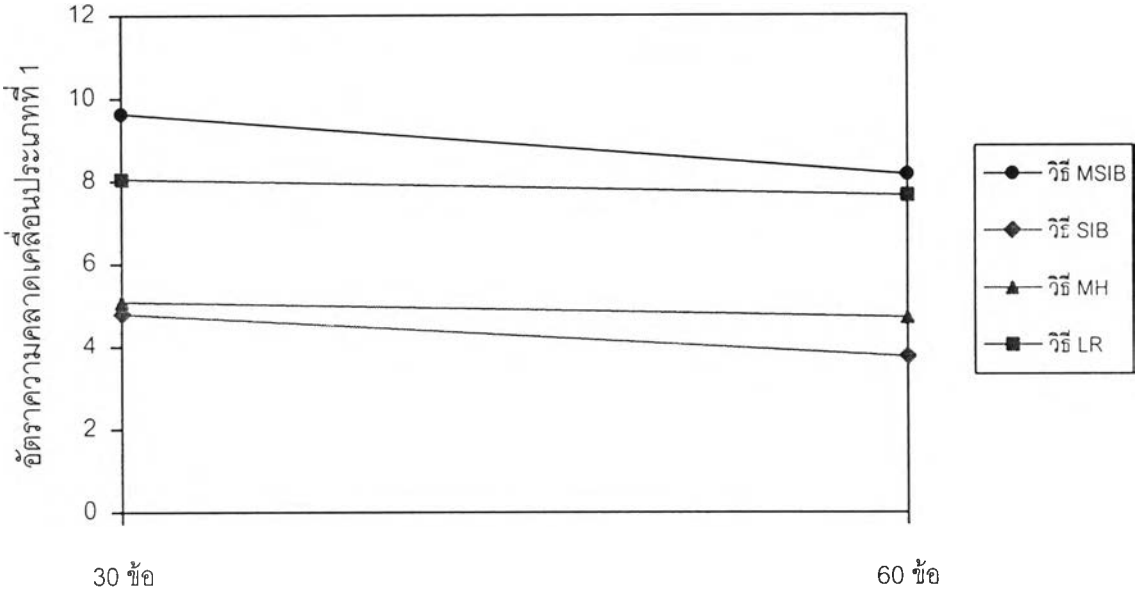
1000 : 250, 1000 : 500 และ 1000 : 1000) ส่วนตัวแปรตาม คือ อัตราความคลาดเคลื่อนประเภทที่ 1 กราฟดังกล่าวแสดงดังภาพที่ 8, 9, 10 และ 11



ภาพที่ 8 กราฟแสดงการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ

จากภาพที่ 8 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันและตัดกัน โดยเส้นกราฟของวิธีซิปเทสท์ปรับใหม่ตัดกับวิธีการถดถอยโลจิสติก และเส้นกราฟของวิธีซิปเทสท์ตัดกับวิธีแมนเทล-แฮนส์เซล แสดงว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและลักษณะของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในลักษณะไม่เป็นลำดับ (disordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบต่างกันจะไม่เหมือนกันภายใต้ทุกลักษณะของข้อสอบ ดังเช่น ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_M; a_L, b_H; a_M, b_L; a_H, b_M และ a_H, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าสูงกว่าวิธีซิปเทสท์ แต่ภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_L อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงกว่าวิธีแมนเทล-แฮนส์เซล เมื่อ

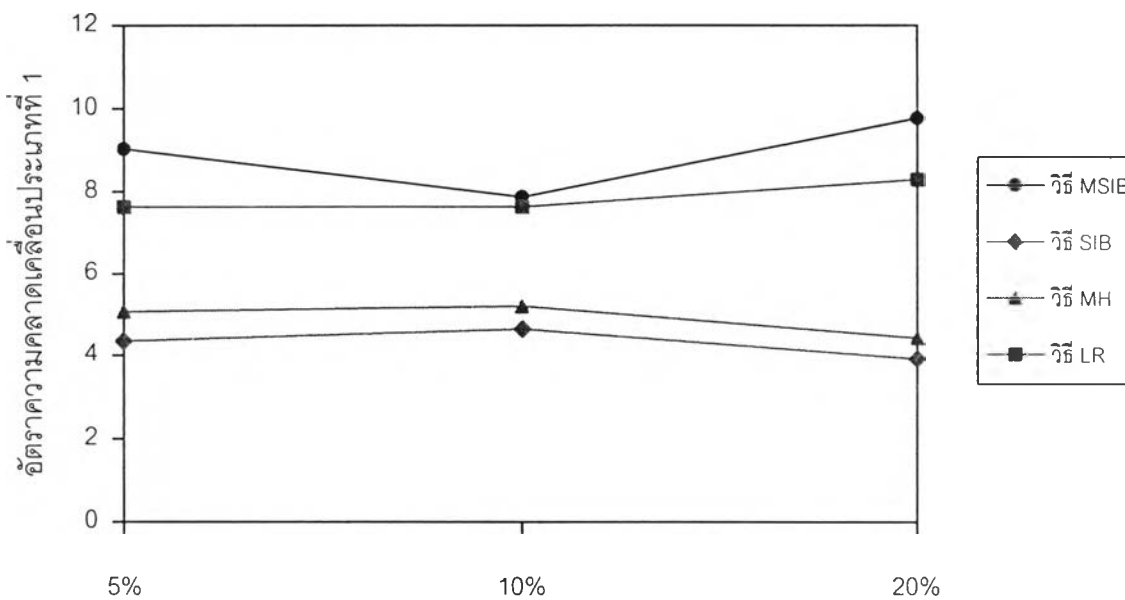
พิจารณาความลาดเอียงของกราฟ ปรากฏว่า เส้นกราฟของทั้ง 4 วิธีมีความลาดเอียงไปในทิศทางเดียวกันเกือบทุกลักษณะของข้อสอบ แสดงว่า ลักษณะของข้อสอบน่าจะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน สรุปได้ว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล



ภาพที่ 9 กราฟแสดงการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ

จากภาพที่ 9 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันแต่ไม่ตัดกัน แสดงว่าปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและความยาวของแบบสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในลักษณะเป็นลำดับ (ordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบต่างกันจะเหมือนกันภายใต้ทุกความยาวของแบบสอบ ดังเช่น ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าทั้ง 3 วิธี ส่วนวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกก็อธิบายได้ในลักษณะเดียวกัน เมื่อพิจารณาความลาดเอียงของกราฟ ปรากฏว่า เส้นกราฟของวิธีซิปเทสท์-

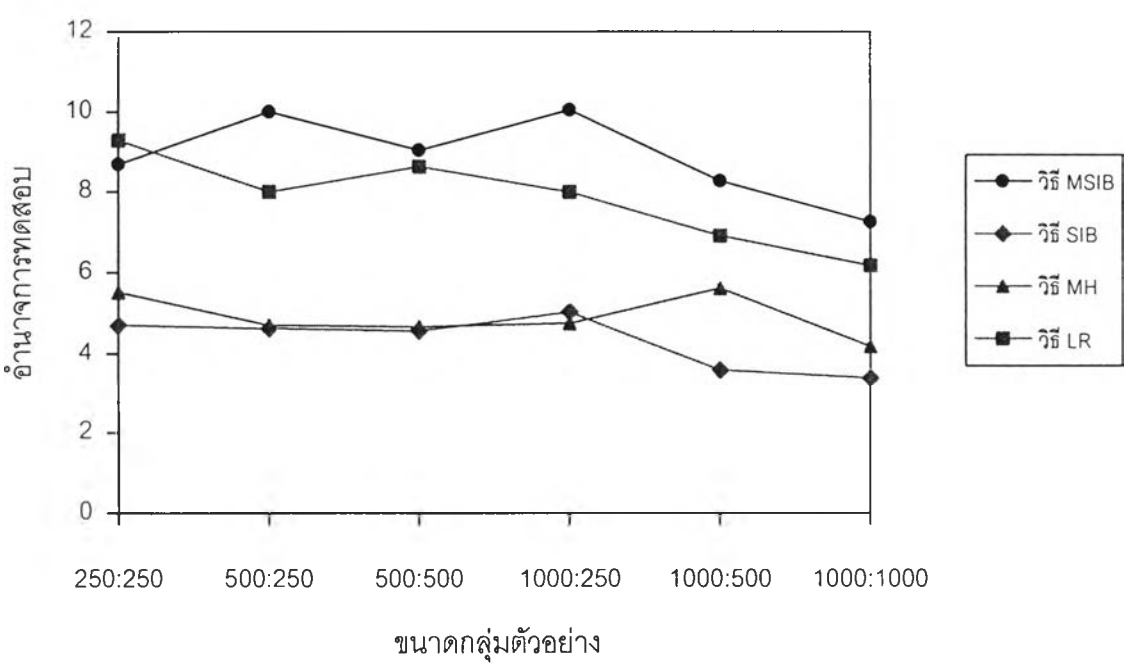
ปรับใหม่และวิธีการถดถอยโลจิสติกมีความลาดเอียงเพียงเล็กน้อย ส่วนเส้นกราฟของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลเกือบจะไม่มีความลาดเอียง โดยขนานกับแกนนอน แสดงว่าความยาวของแบบสอบไม่น่าจะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน สรุปได้ว่า อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล



ภาพที่ 10 กราฟแสดงการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ4 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน มีจำนวนต่างกัน 3 ระดับ

จากภาพที่ 10 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันและไม่ตัดกัน แสดงว่าปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในลักษณะเป็นลำดับ (ordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบต่างกันจะเหมือนกันภายใต้ทุกสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ดังเช่น ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% อัตรา

ความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอีก 3 วิธี สวนวิธีซิปเทสท์ วิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซลก็อธิบายได้ในลักษณะเดียวกัน เมื่อพิจารณาความลาดเอียงของกราฟ ปรากฏว่า เส้นกราฟของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีความลาดเอียงไปในทิศทางเดียวกันในทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ส่วนของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีความลาดเอียงไปในทิศทางเดียวกันในเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ดังนั้นสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันน่าจะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน สรุปได้ว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล



ภาพที่ 11 กราฟแสดงการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ

จากภาพที่ 11 จะเห็นว่าเส้นกราฟของทั้ง 4 วิธีไม่ขนานกันและตัดกัน โดยเส้นกราฟของวิธีซิปเทสท์ปรับใหม่ตัดกับวิธีการถดถอยโลจิสติก และเส้นกราฟของวิธีซิปเทสท์ตัดกับวิธีแมนเทล-แฮนส์เซล แสดงว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปและขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในลักษณะไม่เป็นลำดับ (disordinal interaction effect) นั่นคือ ทิศทางของความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบต่างกันจะไม่เหมือนกันภายใต้ทุกขนาดกลุ่มตัวอย่าง ดังเช่น ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 250 : 250$ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์ปรับใหม่ แต่ภายใต้ขนาดกลุ่มตัวอย่างอื่น ๆ ที่เหลือทั้งหมด อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าวิธีการถดถอยโลจิสติก เมื่อพิจารณาความลาดเอียงของกราฟ พบว่า เส้นกราฟของทั้ง 4 วิธี มีความลาดเอียงไปในทิศทางตรงกันข้าม ดังนั้นขนาดกลุ่มตัวอย่างไม่น่าจะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาเส้นกราฟของทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน สรุปได้ว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล

อย่างไรก็ตาม ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งพิจารณาจากกราฟดังกล่าวเป็นวิธีการเปรียบเทียบอย่างคร่าว ๆ ยังไม่สามารถตอบสมมติฐานได้อย่างแท้จริง ดังนั้นผู้วิจัยจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษา 4 ตัว จากตารางที่ 102 ในภาคผนวก ฉ หน้า 350 – 359 ไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญจะใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ทดสอบในแต่ละระดับของตัวแปรแล้วเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ยกเว้นการเปรียบเทียบภายใต้ความยาวของแบบสอบ 2 ระดับ จะใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (independent sample) โดยจะทดสอบใน 3 กรณีต่อไปนี้ (1) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน (2) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ (3) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษาต่างระดับกัน ผลการวิเคราะห์ดังกล่าวแสดงดังตารางต่อไปนี้

ตารางที่ 36 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	34805.539	1601.846	92.219**
ลักษณะของข้อสอบ (TI)	8	1356.085	169.511	9.759**
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × TI	24	677.718	28.238	1.626*
ความคลาดเคลื่อน	1260	21886.250	17.370	
รวม	1295	58725.592		

**p < .01, *p < .05

จากตารางที่ 36 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และลักษณะของข้อสอบ (TI) มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้ลักษณะของข้อสอบต่างกันจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ดังกล่าวไปเปรียบเทียบความแตกต่าง โดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 37

กรณีที่ 2 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 38

กรณีที่ 3 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 44

ตารางที่ 37 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี MSIB				
ระหว่างกลุ่ม	8	456.556	57.069	2.085*
ภายในกลุ่ม	315	8623.750	27.377	
รวม	323	9080.306		
วิธี SIB				
ระหว่างกลุ่ม	8	186.302	23.288	2.092*
ภายในกลุ่ม	315	3506.833	11.133	
รวม	323	3693.135		
วิธี MH				
ระหว่างกลุ่ม	8	676.500	84.563	6.578**
ภายในกลุ่ม	315	4049.139	12.854	
รวม	323	4725.639		
วิธี LR				
ระหว่างกลุ่ม	8	714.444	89.306	4.930**
ภายในกลุ่ม	315	5706.528	18.116	
รวม	323	6420.972		

**p < .01

จากตารางที่ 37 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ระดับ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 38

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 39

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 40

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ดังนั้นจึงนำอัตราความ-

คลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 41

ตารางที่ 38 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ระดับ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\bar{X}$	9.50	10.64	10.78	9.28	7.36	8.64	7.92	7.56	8.25
a_L, b_L	–	1.14	1.28	0.22	2.14	0.86	1.58	1.94	1.25
a_L, b_M		–	0.14	1.36	3.28	2.00	2.72	3.08	2.39
a_L, b_H			–	1.50	3.42	2.14	2.86	3.22	2.53
a_M, b_L				–	1.92	0.64	1.36	1.72	1.03
a_M, b_M					–	1.28	0.56	0.20	0.89
a_M, b_H						–	0.72	1.08	0.39
a_H, b_L							–	0.36	0.33
a_H, b_M								–	0.69
a_H, b_H									–

*$p < .05$

จากตารางที่ 38 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_H, b_H; a_M, b_L; a_L, b_L; a_H, b_L; a_M, b_H; a_H, b_M และ a_M, b_M ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบต่างกัน 9 ระดับ พบว่า ไม่พบความแตกต่าง นั่นคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ทุกเงื่อนไขของลักษณะของข้อสอบต่างกัน 9 ลักษณะ มีค่าใกล้เคียงกัน

ตารางที่ 39 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\bar{X}$	2.94	5.00	4.28	4.39	4.17	5.64	3.89	3.58	4.89
a_L, b_L	–	2.06	1.34	1.45	1.23	2.70*	0.95	0.64	1.95
a_L, b_M		–	0.72	0.61	0.83	0.64	1.11	1.42	0.11
a_L, b_H			–	0.11	0.11	1.36	0.39	0.70	0.61
a_M, b_L				–	0.22	1.25	0.50	0.81	0.50
a_M, b_M					–	1.47	0.28	0.59	0.72
a_M, b_H						–	1.75	2.06	0.75
a_H, b_L							–	0.31	1.00
a_H, b_M								–	1.31
a_H, b_H									–

*$p < .05$

จากตารางที่ 39 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_H, b_H; a_M, b_L; a_L, b_H; a_M, b_M; a_H, b_L; a_H, b_M และ a_L, b_L ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_L, b_H; a_M, b_L; a_M, b_M; a_H, b_L; a_H, b_M และ a_H, b_H มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ทั้งหมด

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกปานกลางและค่าความยากสูงจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด

ตารางที่ 40 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\bar{X}$	4.53	6.31	7.39	4.92	3.69	5.56	2.31	3.75	5.64
a_L, b_L	–	1.78	2.86*	0.39	0.84	1.03	2.22	0.78	1.11
a_L, b_M		–	1.08	1.39	2.62	0.75	4.00*	2.56	0.67
a_L, b_H			–	2.47	3.70*	1.83	5.08*	3.64*	1.75
a_M, b_L				–	1.23	0.64	2.61	1.17	0.72
a_M, b_M					–	1.87	1.38	0.06	1.95
a_M, b_H						–	3.25*	1.81	0.08
a_H, b_L							–	1.44	3.33*
a_H, b_M								–	1.89
a_H, b_H									–

*$p < .05$

จากตารางที่ 40 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_H, b_H; a_M, b_H; a_M, b_L; a_L, b_L; a_H, b_M; a_M, b_M และ a_H, b_L ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏผลว่า

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L; a_M, b_M; a_H, b_L และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_L กับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ทั้งหมดมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_L, b_H; a_M, b_H และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกต่ำและค่าความยากสูงจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด โดยเฉพาะการตรวจสอบภายใต้เงื่อนไข a_H, b_L พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าประมาณร้อยละ 2.31 และภายใต้เงื่อนไข a_L, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าประมาณร้อยละ 7.39 ดังนั้นในการตรวจสอบภายใต้ 2 เงื่อนไขดังกล่าวจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกันมากที่สุด ประมาณร้อยละ 5.08

ตารางที่ 41 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

TI	a_L, b_L	a_L, b_M	a_L, b_H	a_M, b_L	a_M, b_M	a_M, b_H	a_H, b_L	a_H, b_M	a_H, b_H
$\bar{X}$	9.94	9.25	9.33	8.06	6.19	8.17	6.36	5.36	7.92
a_L, b_L	–	0.69	0.61	1.88	3.75*	1.77	3.58*	4.58*	2.02
a_L, b_M		–	0.08	1.19	3.06	1.08	2.89	3.89*	1.33
a_L, b_H			–	1.27	3.14*	1.16	2.97	3.97*	1.41
a_M, b_L				–	1.87	0.11	1.70	2.70	0.14
a_M, b_M					–	1.98	0.17	0.83	1.73
a_M, b_H						–	1.81	2.81	0.25
a_H, b_L							–	1.00	1.56
a_H, b_M								–	2.56
a_H, b_H									–

*$p < .05$

จากตารางที่ 41 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าสูงสุด รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_L, b_H; a_L, b_M; a_M, b_H; a_M, b_L; a_H, b_H; a_H, b_L; a_M, b_M และ a_H, b_M ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ปรากฏว่า

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_M; a_H, b_L และ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_L; a_M, b_H และ a_H, b_H มีค่าไม่แตกต่างกัน และไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบอื่น ๆ ทั้งหมด

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L และ a_L, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_M และ a_L, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกลดลงจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้น โดยเฉพาะการตรวจสอบภายใต้เงื่อนไข a_H, b_M พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าประมาณร้อยละ 5.36 และภายใต้เงื่อนไข a_L, b_L อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าประมาณร้อยละ 9.94 ดังนั้นการตรวจสอบภายใต้ 2 เงื่อนไขดังกล่าวจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกันมากที่สุดประมาณร้อยละ 4.58

ตารางที่ 42 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบเดียวกัน 9 ลักษณะ

แหล่งของความแปรปรวน	df	SS	MS	F
a_L, b_L				
ระหว่างกลุ่ม	3	1338.688	446.229	17.871**
ภายในกลุ่ม	140	3495.750	24.970	
รวม	143	4834.438		
a_L, b_M				
ระหว่างกลุ่ม	3	728.465	242.822	15.011**
ภายในกลุ่ม	140	2264.694	16.176	
รวม	143	2993.159		

ตารางที่ 42 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
a_L, b_H				
ระหว่างกลุ่ม	3	853.556	284.519	13.269**
ภายในกลุ่ม	140	3002.000	21.443	
รวม	143	3855.556		
a_M, b_L				
ระหว่างกลุ่ม	3	611.910	203.970	12.424**
ภายในกลุ่ม	140	2298.417	16.417	
รวม	143	2910.327		
a_M, b_M				
ระหว่างกลุ่ม	3	320.354	106.785	7.610**
ภายในกลุ่ม	140	1964.583	14.033	
รวม	143	2284.937		
a_M, b_H				
ระหว่างกลุ่ม	3	287.500	95.833	5.147**
ภายในกลุ่ม	140	2606.500	18.618	
รวม	143	2894.000		
a_H, b_L				
ระหว่างกลุ่ม	3	676.743	225.581	16.779**
ภายในกลุ่ม	140	1882.250	13.445	
รวม	143	2558.993		

ตารางที่ 42 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
a_H, b_M				
ระหว่างกลุ่ม	3	367.743	122.581	9.263**
ภายในกลุ่ม	140	1852.694	13.234	
รวม	143	2220.437		
a_H, b_H				
ระหว่างกลุ่ม	3	298.299	99.433	5.525**
ภายในกลุ่ม	140	2519.361	17.995	
รวม	143	2817.660		

**p < .01

จากตารางที่ 42 พบว่า ภายใต้ลักษณะของข้อสอบ 9 ลักษณะ คือ a_L, b_L; a_L, b_M; a_L, b_H; a_M, b_L; a_M, b_M; a_M, b_H; a_H, b_L; a_H, b_M และ a_H, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้ลักษณะของข้อสอบทั้ง 9 ลักษณะไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 46

ตารางที่ 43 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบเดียวกัน 9 ลักษณะ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\overline{X}$	9.50	2.94	4.53	9.94
a_L, b_L	วิธี MSIB	–	6.56*	4.97*	0.44
	วิธี SIB		–	1.59	3.38*
	วิธี MH			–	5.41*
	วิธี LR				–
	$\overline{X}$	10.64	5.00	6.31	9.25
a_L, b_M	วิธี MSIB	–	5.64*	4.33*	1.39
	วิธี SIB		–	1.31	4.25*
	วิธี MH			–	2.94*
	วิธี LR				–
	$\overline{X}$	10.78	4.28	7.39	9.33
a_L, b_H	วิธี MSIB	–	6.50*	3.39*	1.45
	วิธี SIB		–	3.11*	5.05*
	วิธี MH			–	1.94
	วิธี LR				–
	$\overline{X}$	9.28	4.39	4.92	8.06
a_M, b_L	วิธี MSIB	–	4.89*	4.36*	1.22
	วิธี SIB		–	0.53	3.67*
	วิธี MH			–	3.14*
	วิธี LR				–

ตารางที่ 43 (ต่อ)

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
$\overline{X}$		7.36	4.17	3.69	6.19
a_M, b_M	วิธี MSIB	–	3.19*	3.67*	1.17
	วิธี SIB		–	0.48	0.02
	วิธี MH			–	2.50*
	วิธี LR				–
$\overline{X}$		8.64	5.64	5.56	8.17
a_M, b_H	วิธี MSIB	–	3.00*	3.08*	0.47
	วิธี SIB		–	0.08	2.53
	วิธี MH			–	2.61
	วิธี LR				–
$\overline{X}$		8.78	3.89	2.31	6.36
a_H, b_L	วิธี MSIB	–	4.89*	6.47*	2.42
	วิธี SIB		–	1.58	2.47*
	วิธี MH			–	4.05*
	วิธี LR				–
$\overline{X}$		7.56	3.58	3.75	5.36
a_H, b_M	วิธี MSIB	–	3.98*	3.81*	2.20
	วิธี SIB		–	0.17	1.78
	วิธี MH			–	1.61
	วิธี LR				–

ตารางที่ 43 (ต่อ)

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
$\bar{X}$		8.25	4.89	5.64	7.92
a_H, b_H	วิธี MSIB	–	3.36*	2.61*	0.33
	วิธี SIB		–	0.75	3.03*
	วิธี MH			–	2.28
	วิธี LR				–

*p < .05

จากตารางที่ 43 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยลักษณะของข้อสอบเดียวกัน 9 ลักษณะ ปรากฏผลดังนี้

ภายใต้ลักษณะของข้อสอบ a_L, b_L อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_L, b_M และ a_M, b_L ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_L, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีซิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_M อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_L อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความ-

คลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_M อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซลและซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซลและซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลการศึกษาดังกล่าว แสดงว่า ภายใต้ลักษณะของข้อสอบ 9 ลักษณะ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภท 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไขของลักษณะของข้อสอบ

ตารางที่ 44 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ

วิธีการตรวจสอบ DIF (TI)	MSIB (a_L, b_L)	MSIB (a_L, b_M)	MSIB (a_L, b_H)	MSIB (a_M, b_L)	MSIB (a_M, b_M)	MSIB (a_M, b_H)	MSIB (a_H, b_L)	MSIB (a_H, b_M)	MSIB (a_H, b_H)	SIB (a_L, b_L)	SIB (a_L, b_M)	SIB (a_L, b_H)	SIB (a_M, b_L)	SIB (a_M, b_M)	SIB (a_M, b_H)	SIB (a_H, b_L)	SIB (a_H, b_M)	SIB (a_H, b_H)
$\bar{X}$	9.50	10.64	10.78	9.28	7.36	8.64	7.92	7.56	8.25	2.94	5.00	4.28	4.39	4.17	5.64	3.89	3.58	4.89
MSIB (a_L, b_L)	–	1.14	1.28	0.22	2.14	0.86	1.58	1.94	1.25	6.56*	4.50*	5.22*	5.11*	5.33*	3.86*	5.61*	5.92*	4.61*
MSIB (a_L, b_M)		–	0.14	1.36	3.28	2.00	2.72	3.08	2.39	7.70*	5.64*	6.36*	6.25*	6.47*	5.00*	6.75*	7.06*	5.75*
MSIB (a_L, b_H)			–	1.50	3.42	2.14	2.86	3.22	2.53	7.84*	5.78*	6.50*	6.39*	6.61*	5.14*	6.89*	7.20*	5.89*
MSIB (a_M, b_L)				–	1.92	0.64	1.36	1.72	1.03	6.34*	4.28*	5.00*	4.89*	5.11*	3.64	5.39*	5.70*	4.39*
MSIB (a_M, b_M)					–	1.28	0.56	0.20	0.89	4.42*	2.36	3.08	2.97	3.19	1.72	3.47	3.78*	2.47
MSIB (a_M, b_H)						–	0.72	1.08	0.39	5.70*	3.64	4.36*	4.25*	4.47*	3.00	4.75*	5.06*	3.75
MSIB (a_H, b_L)							–	0.36	0.33	4.98*	2.92	3.64	3.53	3.75	2.28	4.03*	4.34*	3.03*
MSIB (a_H, b_M)								–	0.69	4.62*	2.56	3.28	3.17	3.39	1.92	3.67	3.98*	2.67
MSIB (a_H, b_H)									–	5.31*	3.25	3.97*	3.86*	4.08*	2.61	4.36*	4.67*	3.36
SIB (a_L, b_L)										–	2.06	1.34	1.45	1.23	2.70	0.95	0.64	1.95
SIB (a_L, b_M)											–	0.72	0.61	0.83	0.64	1.11	1.42	0.11
SIB (a_L, b_H)												–	0.11	0.11	1.36	0.39	0.70	0.61
SIB (a_M, b_L)													–	0.22	1.25	1.50	0.81	0.50
SIB (a_M, b_M)														–	1.47	0.28	0.59	0.72
SIB (a_M, b_H)															–	1.75	2.06	0.75
SIB (a_H, b_L)																–	0.31	1.00
SIB (a_H, b_M)																	–	1.31
SIB (a_H, b_H)																		–

ตารางที่ 44 (ต่อ)

วิธีการตรวจสอบ DIF (TI)	MH (a_L, b_L)	MH (a_L, b_M)	MH (a_L, b_H)	MH (a_M, b_L)	MH (a_M, b_M)	MH (a_M, b_H)	MH (a_H, b_L)	MH (a_H, b_M)	MH (a_H, b_H)	LR (a_L, b_L)	LR (a_L, b_M)	LR (a_L, b_H)	LR (a_M, b_L)	LR (a_M, b_M)	LR (a_M, b_H)	LR (a_H, b_L)	LR (a_H, b_M)	LR (a_H, b_H)
$\bar{X}$	4.53	6.31	7.39	4.92	3.69	5.56	2.31	3.75	5.64	9.94	9.26	9.33	8.06	6.19	8.17	6.36	5.36	7.92
MSIB (a_L, b_L)	4.97*	3.19	2.11	4.58*	5.81*	3.94*	7.19*	5.75*	3.86	0.44	0.24	0.17	1.44	3.31	1.33	3.14	4.14*	1.58
MSIB (a_L, b_M)	6.11*	4.33*	3.25	5.72*	6.95*	5.08*	8.33*	6.89*	5.00*	0.70	1.38	1.31	2.58	4.45*	2.47	4.28*	5.28*	2.72
MSIB (a_L, b_H)	6.25*	4.47*	3.39	5.86*	7.09*	5.22*	8.47*	7.03*	5.14*	0.84	1.52	1.45	2.72	4.59*	2.61	4.42*	5.42*	8.91*
MSIB (a_M, b_L)	4.75*	2.97	1.89	4.36*	5.59*	3.72	6.97*	5.53*	3.64	0.66	0.02	0.05	1.22	3.09	1.11	2.92	3.92*	1.36
MSIB (a_M, b_M)	2.83	1.05	0.03	2.44	3.67	1.80	5.05*	3.61	1.72	2.58	1.90	1.97	0.70	1.17	0.81	1.00	2.00	0.56
MSIB (a_M, b_H)	4.11*	2.33	1.25	3.72	4.95*	3.08	6.33*	4.89*	3.00	1.30	0.62	0.69	0.58	2..45	0.47	2.28	3.28	0.72
MSIB (a_H, b_L)	3.39	1.61	0.53	3.00	4.23*	2.36	5.61*	4.17*	2.28	2.02	1.34	1.41	0.14	1.73	0.25	1.56	2.56	0.00
MSIB (a_H, b_M)	3.03	1.25	0.17	2.64	3.87*	2.00	5.25*	3.81*	1.92	2.38	1.70	1.77	0.50	1.37	0.61	1.20	2.20	0.36
MSIB (a_H, b_H)	3.72	1.94	0.86	3.33	4.56*	2.69	5.94*	4.50*	2.61	1.69	1.01	1.08	0.19	2.06	0.08	1.89	2.89	0.33
SIB (a_L, b_L)	1.59	3.37	4.45*	1.98	0.75	2.62	0.63	0 81	2.70	7.00*	6.32*	6.39*	5.12*	3.25	5.23*	3.42	2.42	4.98*
SIB (a_L, b_M)	0.47	1.31	2.39	0.08	1.31	0.56	2.69	1.25	0.64	4.94*	4.26*	4.33*	3.06	1.19	3.17	1.36	0.36	2.92
SIB (a_L, b_H)	0.25	2.03	3.11	0.64	0.59	1.28	1.97	0.53	1.36	5.66*	4.98*	5.05*	3.78*	1.91	3.89*	2.08	1.08	3.64
SIB (a_M, b_L)	0.14	1.92	3.00	0.53	0.70	1.17	2.08	0.64	1.25	5.55*	4.87*	4.94*	3.67	1.81	3.78*	1.97	0.97	3.53
SIB (a_M, b_M)	0.36	2.14	3.22	0.75	0.48	1.39	1.86	0.42	1.47	5.77*	5.09*	5.16*	3.89*	2.02	4.00*	2.19	1.19	3.75
SIB (a_M, b_H)	1.11	0.67	1.75	0.72	1.95	0.08	3.33	1.89	0.00	4.30*	3.62	3.69	2.42	0.55	2.53	0.72	0.28	2.28
SIB (a_H, b_L)	0.64	2.42	3.50	1.89	0.20	1.67	1.58	0.14	1.75	6.10*	5.37*	5.44*	4.17*	2.30	4.28*	2.47	1.47	4.03*
SIB (a_H, b_M)	0.95	2.73	3.81*	1.34	0.11	1.98	1.27	0.17	2.06	6.36*	5.68*	5.75*	4.48*	2.61	4.59*	2.78	1.78	4.34*
SIB (a_H, b_H)	0.36	1.42	2.50	0.03	1.20	0.67	2.58	1.14	0.75	5.05*	4.37*	4.44*	3.17	1.30	3.28	1.47	0.47	3.03

ตารางที่ 44 (ต่อ)

วิธีการตรวจสอบ	MH	MH	MH	MH	MH	MH	MH	MH	MH	LR	LR	LR	LR	LR	LR	LR	LR	LR
DIF (TI)	(a_L, b_L)	(a_L, b_M)	(a_L, b_H)	(a_M, b_L)	(a_M, b_M)	(a_M, b_H)	(a_H, b_L)	(a_H, b_M)	(a_H, b_H)	(a_L, b_L)	(a_L, b_M)	(a_L, b_H)	(a_M, b_L)	(a_M, b_M)	(a_M, b_H)	(a_H, b_L)	(a_H, b_M)	(a_H, b_H)
$\bar{X}$	4.53	6.31	7.39	4.92	3.69	5.56	2.31	3.75	5.64	9.94	9.26	9.33	8.06	6.19	8.17	6.36	5.36	7.92
MH (a_L, b_L)	–	1.78	2.86	0.39	0.84	1.03	2.22	0.78	1.11	5.41*	4.73*	4.80*	3.53	1.66	3.64	1.83	0.83	3.39
MH (a_L, b_M)		–	1.08	1.39	2.62	0.75	4.00*	2.56	0.67	3.63	2.95	3.02	1.75	0.12	1.86	0.05	0.95	1.61
MH (a_L, b_H)			–	2.47	3.70	1.83	5.08*	3.64	1.75	2.55	1.87	1.94	0.67	1.20	0.78	0.83	2.03	0.53
MH (a_M, b_L)				–	2.46	0.64	2.61	1.17	0.72	5.02*	4.34*	4.41*	3.14	1.27	3.25	1.44	0.44	3.00
MH (a_M, b_M)					–	1.87	1.38	0.06	1.95	6.25*	5.57*	5.64*	4.37*	2.50	4.48*	2.67	1.67	4.23*
MH (a_M, b_H)						–	3.25	1.81	0.08	4.38*	3.70	3.77	2.50	0.63	2.61	0.80	0.20	2.36
MH (a_H, b_L)							–	1.44	3.33	7.63*	6.95*	7.02*	5.75*	3.88*	5.86*	4.05*	3.05	5.61*
MH (a_H, b_M)								–	1.89	6.19*	5.51*	5.58*	4.31*	2.44	4.42*	2.61	1.61	4.17*
MH (a_H, b_H)									–	4.30*	3.62	3.69	2.42	0.55	2.53	0.72	0.28	2.28
LR (a_L, b_L)										–	0.68	0.61	1.88	3.75	1.77	3.58	4.58*	0.02
LR (a_L, b_M)											–	0.07	1.20	3.07	1.09	2.90	3.90*	1.34
LR (a_L, b_H)												–	1.27	3.14	1.16	2.97	3.97*	1.41
LR (a_M, b_L)													–	1.87	0.11	1.70	2.70	0.14
LR (a_M, b_M)														–	1.98	0.17	0.83	1.73
LR (a_M, b_H)															–	1.81	2.81	0.25
LR (a_H, b_L)																–	1.00	1.56
LR (a_H, b_M)																	–	2.56
LR (a_H, b_H)																		–

*$p < .05$

จากตารางที่ 44 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งมีทั้งหมด 36 ค่า เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB (a_L,b_H)มีค่าสูงสุด รองลงมาได้แก่ วิธีMSIB(a_L,b_M)และวิธีLR(a_L,b_L) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMH(a_H,b_L)มีค่าต่ำสุด รองลงมาได้แก่ วิธีSIB(a_L,b_L) และวิธีSIB(a_H,b_M) ตามลำดับ ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบดังกล่าวกับวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ปรากฏผล ดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_H) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB(a_H,b_H), วิธีMH(a_L,b_H), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_L) และวิธีLR(a_M,b_H)

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_M) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L), วิธีMSIB(a_L,b_H), วิธีMSIB(a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB(a_H,b_H), วิธีMH(a_L,b_H), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_L), วิธีLR(a_M,b_H) และวิธีLR(a_H,b_H) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีLR(a_L,b_L) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB(a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB(a_H,b_H), วิธีMH(a_L,b_M), วิธีMH(a_L,b_H), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_L), วิธีLR(a_M,b_M), วิธีLR(a_M,b_H), วิธีLR(a_H,b_L) และวิธีLR(a_H,b_H) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB(a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB(a_H,b_H), วิธีMH(a_L,b_M), วิธีMH(a_L,b_H), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_L), วิธีLR(a_M,b_M), วิธีLR(a_M,b_H), วิธีLR(a_H,b_L) และวิธีLR(a_H,b_H) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMH(a_H,b_L) พบว่า มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB (a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB (a_H,b_H), วิธีMH (a_L,b_M), วิธีMH(a_L,b_H), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR (a_M,b_L), วิธีLR(a_M,b_M), วิธีLR(a_M,b_H), วิธีLR(a_H,b_L) และวิธีLR(a_H,b_H) แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(a_L,b_L) พบว่า มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB (a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB (a_H,b_H), วิธีMH(a_L,b_H), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_L), วิธีLR (a_M,b_H) และวิธีLR(a_H,b_H) แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(a_H,b_M) พบว่า มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(a_L,b_L), วิธีMSIB(a_L,b_M), วิธีMSIB(a_L,b_H), วิธีMSIB (a_M,b_L), วิธีMSIB(a_M,b_M), วิธีMSIB(a_M,b_H), วิธีMSIB(a_H,b_L), วิธีMSIB(a_H,b_M), วิธีMSIB (a_H,b_H), วิธีMH(a_L,b_H), วิธีLR(a_L,b_L), วิธีLR(a_L,b_M), วิธีLR(a_L,b_H), วิธีLR(a_M,b_L), วิธีLR (a_M,b_H) และวิธีLR(a_H,b_H) แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ลักษณะของข้อสอบอื่น ๆ ที่เหลือ

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป เมื่อใช้วิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a ต่ำกับ b สูงจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลเกือบทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะ

ของข้อสอบ a สูงกับ b ต่ำจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ

ตารางที่ 45 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	4805.539	1601.846	87.323**
ความยาวของแบบสอบ (TL)	8	219.204	219.204	11.950**
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × TL	24	73.743	24.581	1.340
ความคลาดเคลื่อน	1260	23627.105	18.344	
รวม	1295	28725.591		

**p < .01

จากตารางที่ 45 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) ปรากฏว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยาวของแบบสอบ (TL) ไม่มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้ความยาวของแบบสอบต่างกันแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของความยาวของแบบสอบ (TL) ปรากฏว่า ความยาวของแบบสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ความยาวของแบบสอบต่างกันจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4

วิธีภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับไปเปรียบเทียบโดยใช้สถิติ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (independent sample) ผลการเปรียบเทียบแสดงดังตารางที่ 46

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบ DIF มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบ 2 ระดับไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 47

ตารางที่ 46 ผลการวิเคราะห์ t-test ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ระดับ

TL	N	$\bar{X}$	SD	t-test
วิธี MSIB				
30 ข้อ	162	9.62	6.03	2.55**
60 ข้อ	162	8.14	4.36	
วิธี SIB				
30 ข้อ	162	4.84	4.06	2.86**
60 ข้อ	162	3.78	2.43	
วิธี MH				
30 ข้อ	162	5.07	4.56	0.80
60 ข้อ	162	4.73	2.92	
วิธี LR				
30 ข้อ	162	8.04	5.06	0.81
60 ข้อ	162	7.64	3.76	

*p < .01

จากตารางที่ 46 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ระดับ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้ออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าสูงกว่าอัตราความ-คลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้ออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก แต่จะมีผลต่ออัตราความ-คลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบจาก 30 ข้อ เป็น 60 ข้อจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธี มีค่าลดลงประมาณร้อยละ 1.48 และ 1.06 ตามลำดับ

ตารางที่ 47 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบระดับเดียวกัน 2 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
30 ข้อ				
ระหว่างกลุ่ม	3	2644.844	881.615	35.545**
ภายในกลุ่ม	644	15972.809	24.802	
รวม	647	18617.653		
60 ข้อ				
ระหว่างกลุ่ม	3	2234.438	744.813	62.665**
ภายในกลุ่ม	644	7654.296	11.886	
รวม	647	9888.734		

**p < .01

จากตารางที่ 47 ปรากฏว่า ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้ความยาวของแบบสอบทั้ง 2 ระดับไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 48

ตารางที่ 48 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบระดับเดียวกัน 2 ระดับ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\bar{X}$	9.62	4.84	5.07	8.04
30 ข้อ	วิธี MSIB	–	4.78*	4.55*	1.58*
	วิธี SIB		–	0.23	3.20*
	วิธี MH			–	2.97*
	วิธี LR				–
	$\bar{X}$	8.14	3.78	4.73	7.64
60 ข้อ	วิธี MSIB	–	4.36*	3.41*	0.50
	วิธี SIB		–	0.95	3.86*
	วิธี MH			–	2.91*
	วิธี LR				–

*p < .05

จากตารางที่ 48 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ ปรากฏผลดังนี้

ภายใต้ความยาวของแบบสอบ 30 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่

ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ความยาวของแบบสอบ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไขของความยาวของแบบสอบ

ตารางที่ 49 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	4805.539	1601.846	87.030**
สัดส่วนของ DIF (%DIF)	2	16.673	8.336	.453
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × %DIF	6	270.426	45.071	2.449*
ความคลาดเคลื่อน	1284	23632.954	18.406	
รวม	1295	28725.592		

**p < .01, *p < .05

จากตารางที่ 49 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) ปรากฏว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 มาเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 50

กรณีที่ 2 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 52

กรณีที่ 3 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 54

ตารางที่ 50 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี MSIB				
ระหว่างกลุ่ม	2	193.907	96.954	3.502*
ภายในกลุ่ม	321	8886.398	27.683	
รวม	323	9080.305		

ตารางที่ 50 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี SIB				
ระหว่างกลุ่ม		27.043	13.522	1.184
ภายในกลุ่ม		3666.093	11.421	
รวม		3693.136		
วิธี MH				
ระหว่างกลุ่ม		34.130	17.065	1.168
ภายในกลุ่ม		4691.509	14.615	
รวม		4725.639		
วิธี LR				
ระหว่างกลุ่ม		32.019	16.009	.804
ภายในกลุ่ม		6388.954	19.903	
รวม		6420.973		

**p < .01

จากตารางที่ 50 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 51

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ มีค่าไม่แตกต่างกัน แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ มีค่าไม่แตกต่างกัน แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ มีค่าไม่แตกต่างกัน แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก

ตารางที่ 51 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Tukey

%DIF	5%	10%	20%
$\bar{X}$	9.02	7.87	9.75
5%	–	1.15	0.73
10%		–	1.88*
20%			–

*p < .05

จากตารางที่ 51 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบจำนวน 20% มีค่าสูงสุด รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่า

สูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าใกล้เคียงกับภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% แต่ถ้าเพิ่มสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันจาก 10% เป็น 20% จะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้นสูงสุด ประมาณ ร้อยละ 1.88

ตารางที่ 52 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
5%				
ระหว่างกลุ่ม	3	1539.137	513.046	29.298**
ภายในกลุ่ม	428	7494.806	17.511	
รวม	431	9033.942		
10%				
ระหว่างกลุ่ม	3	889.081	296.360	17.237**
ภายในกลุ่ม	428	7358.583	17.193	
รวม	431	8247.664		
20%				
ระหว่างกลุ่ม	3	2647.748	882.583	43.026**
ภายในกลุ่ม	428	8779.565	20.513	
รวม	431	11427.312		

**p < .01

จากตารางที่ 52 พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ คือ 5%, 10% และ 20% อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 ระดับ ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 53

ตารางที่ 53 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\bar{X}$	9.02	4.35	5.06	7.61
5%	วิธี MSIB	–	4.67*	3.96*	1.41
	วิธี SIB		–	0.71	3.26*
	วิธี MH			–	2.55*
	วิธี LR				–
	$\bar{X}$.	7.87	4.64	5.19	7.63
10%	วิธี MSIB	–	3.23*	2.68*	0.24
	วิธี SIB		–	0.55	2.99*
	วิธี MH			–	2.44*
	วิธี LR				–
	$\bar{X}$	9.75	3.94	4.44	8.29
20%	วิธี MSIB	–	5.81*	5.31*	1.46
	วิธี SIB		–	0.50	4.35*
	วิธี MH			–	3.85*
	วิธี LR				–

*$p < .05$

จากตารางที่ 53 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ ปรากฏผลดังนี้

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้ทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 54 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ

วิธีการตรวจสอบ DIF (TI)	MSIB (5%)	MSIB (10%)	MSIB (20%)	SIB (5%)	SIB (10%)	SIB (20%)	MH (5%)	MH (10%)	MH (20%)	LR (5%)	LR (10%)	LR (20%)
$\overline{X}$	9.02	7.87	9.75	4.35	4.64	3.94	5.06	5.19	4.44	7.61	7.63	8.29
MSIB (5%)	–	1.15	0.73	4.67*	4.38*	5.08*	3.96*	3.83*	4.58*	1.41	1.39	0.73
MSIB (10%)		–	1.88	3.52*	3.23*	3.93*	2.81*	2.68*	3.43*	0.26	0.24	0.42
MSIB (20%)			–	5.40*	5.11*	5.81*	4.69*	4.56*	5.31*	2.14*	2.12*	1.46
SIB (5%)				–	0.29	0.41	0.71	0.84	0.09	3.26*	3.28*	3.94*
SIB (10%)					–	0.70	0.42	0.55	0.20	2.97*	2.99*	3.65*
SIB (20%)						–	1.12	1.25	0.50	3.67*	3.69*	4.35*
MH (5%)							–	0.13	0.62	2.55*	2.57*	3.23*
MH (10%)								–	0.75	2.42*	2.44*	3.10*
MH (20%)									–	3.17*	3.19*	3.85*
LR (5%)										–	0.02	0.68
LR (10%)											–	0.66
LR (20%)												–

*p < .05

จากตารางที่ 54 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ โดยเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภท 1 ซึ่งมีทั้งหมด 12 ค่า เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(20%)มีค่าสูงสุด รองลงมาได้แก่ วิธีMSIB(5%) และวิธีLR(20%) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(20%)มีค่าต่ำสุด รองลงมาได้แก่ วิธีSIB(5%) และวิธีMH(20%) ตามลำดับ ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันดังกล่าวกับวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ปรากฏผล ดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(20%) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(5%), วิธีMSIB(10%) และวิธีLR(20%) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(5%) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(10%), วิธีMSIB(20%), วิธีLR(5%), วิธีLR(10%) และวิธีLR(20%) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีLR(20%) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMSIB(5%), วิธีMSIB(10%), วิธีMSIB(20%), วิธีLR(5%) และวิธีLR(10%) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(20%) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(5%), วิธีSIB(10%), วิธีMH(5%), วิธีMH(10%) และวิธีMH(20%) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(5%) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(10%), วิธีSIB(20%), วิธีMH(5%), วิธีMH(10%), และวิธีMH(20%) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีMH(20%) พบว่า มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีSIB(5%), วิธีSIB(10%), วิธีSIB(20%), วิธีMH(5%) และวิธีMH(10%) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป เมื่อใช้วิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ ในขณะเดียวกัน เมื่อใช้วิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกในทุกเงื่อนไขของการตรวจสอบ

ตารางที่ 55 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทางของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
ผลหลัก				
วิธีการตรวจสอบ DIF	3	4805.539	1601.846	88.305**
ขนาดกลุ่มตัวอย่าง (N)	5	512.291	102.458	5.648**
ปฏิสัมพันธ์ 2 ทาง				
วิธีการตรวจสอบ DIF × N	15	333.817	22.254	1.227
ความคลาดเคลื่อน	1272	23073.944	18.140	
รวม	1295	28725.592		

**p < .01

จากตารางที่ 55 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two-way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และขนาดกลุ่มตัวอย่าง (N) ไม่มีนัยสำคัญทางสถิติ แสดงว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันภายใต้ขนาดกลุ่มตัวอย่างต่างกันแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและขนาดกลุ่มตัวอย่างไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของปัจจัยขนาดกลุ่มตัวอย่าง (N) พบว่า ขนาดกลุ่มตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ขนาดกลุ่มตัวอย่างต่างกันจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับไปเปรียบเทียบความแตกต่าง โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 56

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบ DIF มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการตรวจสอบต่างกันจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ผลการเปรียบเทียบแสดงดังตารางที่ 57

ตารางที่ 56 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี MSIB				
ระหว่างกลุ่ม	5	304.102	60.820	2.204
ภายในกลุ่ม	318	8776.204	27.598	
รวม	323	9080.306		

ตารางที่ 56 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
วิธี SIB				
ระหว่างกลุ่ม	5	116.469	23.294	2.071
ภายในกลุ่ม	318	3576.667	11.247	
รวม	323	3693.136		
วิธี MH				
ระหว่างกลุ่ม	8	83.806	16.761	1.148
ภายในกลุ่ม	315	4641.833	14.597	
รวม	323	4725.639		
วิธี LR				
ระหว่างกลุ่ม	5	341.731	68.346	3.575**
ภายในกลุ่ม	318	6079.241	19.117	
รวม	323	6420.972		

**p < .01

จากตารางที่ 56 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าไม่แตกต่างกัน แสดงว่า ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์-ปรับใหม่

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าไม่แตกต่างกัน แสดงว่า ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับมีค่าไม่แตกต่างกัน แสดงว่า ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างในระดับอื่น ๆ ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 57

ตารางที่ 57 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

$N_R : N_F$	250 : 250	500 : 250	500 : 500	1000 : 250	1000 : 500	1000 : 1000
$\bar{X}$	9.28	8.09	8.63	7.94	6.93	6.19
250 : 250	–	1.19	0.65	1.34	2.35	3.09*
500 : 250		–	0.54	0.15	1.16	1.90
500 : 500			–	0.69	1.70	2.44*
1000 : 250				–	1.01	1.75
1000 : 500					–	0.74
1000 :1000						–

*$p < .05$

จากตารางที่ 57 ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 500, 1000 : 250, 500 : 250, 1000 : 500 และ 1000 : 1000 ตามลำดับ เมื่อเปรียบเทียบอัตรา

ความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 และ 500 : 500 มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000 มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก $N_R : N_F$ = 250 : 250 เป็น $N_R : N_F$ = 1000 : 1000 จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้นสูงสุด ประมาณร้อยละ 3.09

ตารางที่ 58 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F
$N_R : N_F$ = 250 : 250				
ระหว่างกลุ่ม	3	841.014	280.338	16.963**
ภายในกลุ่ม	212	3503.611	16.526	
รวม	215	4344.625		
$N_R : N_F$ = 500 : 250				
ระหว่างกลุ่ม	3	1133.199	377.733	19.096**
ภายในกลุ่ม	212	4193.463	19.780	
รวม	215	5326.662		

ตารางที่ 58 (ต่อ)

แหล่งของความแปรปรวน	df	SS	MS	F
$N_R : N_F$ = 500 : 500				
ระหว่างกลุ่ม	3	967.481	322.494	15.696**
ภายในกลุ่ม	212	4355.852	20.546	
รวม	215	5323.333	347.093	
$N_R : N_F$ = 1000 : 250				
ระหว่างกลุ่ม	3	1041.278	16.907	16.907**
ภายในกลุ่ม	212	4352.259	20.530	
รวม	215	5393.537		
$N_R : N_F$ = 1000 : 500				
ระหว่างกลุ่ม	3	636.704	212.235	12.872**
ภายในกลุ่ม	212	3495.444	16.488	
รวม	215	4132.148		
$N_R : N_F$ = 1000 : 1000				
ระหว่างกลุ่ม	3	840.056	280.019	15.419**
ภายในกลุ่ม	212	3850.148	18.161	
รวม	215	4690.204		

**p < .01

จากตารางที่ 58 พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 6 ระดับ คือ $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำผลการทดสอบภายใต้ขนาดกลุ่มตัวอย่างทั้ง 6 ระดับ

ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลการเปรียบเทียบแสดงดังตารางที่ 59

ตารางที่ 59 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ โดยใช้วิธีการทดสอบของ Tukey

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\overline{X}$	8.69	4.69	5.52	9.28
250 : 250	วิธี MSIB	–	4.20*	3.37*	0.59
	วิธี SIB		–	0.83	4.59*
	วิธี MH			–	3.76*
	วิธี LR				–
	$\overline{X}$	10.00	4.61	4.69	8.09
500 : 250	วิธี MSIB	–	5.39*	5.31*	1.91
	วิธี SIB		–	0.08	3.48*
	วิธี MH			–	3.40*
	วิธี LR				–
	$\overline{X}$	9.04	4.56	4.67	8.63
500 : 500	วิธี MSIB	–	4.48*	4.37*	0.41
	วิธี SIB		–	0.11	4.07*
	วิธี MH			–	3.96*
	วิธี LR				–
	$\overline{X}$	10.04	5.02	4.74	7.94
1000 : 250	วิธี MSIB	–	5.02*	5.30*	2.10
	วิธี SIB		–	0.28	2.92*
	วิธี MH			–	3.20*
	วิธี LR				–

ตารางที่ 59 (ต่อ)

วิธีการตรวจสอบ DIF		วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR
	$\bar{X}$	8.24	3.59	5.61	6.93
1000 : 500	วิธี MSIB	–	4.65*	2.63*	1.31
	วิธี SIB		–	2.02*	3.34*
	วิธี MH			–	1.32
	วิธี LR				–
	$\bar{X}$	7.28	3.39	4.17	6.19
1000 : 1000	วิธี MSIB	–	3.89*	3.11*	1.09
	วิธี SIB		–	0.78	2.80*
	วิธี MH			–	2.02*
	วิธี LR				–

*p < .05

จากตารางที่ 59 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ ปรากฏผลดังนี้

ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 500 : 250, 500 : 500 และ 1000 : 1000 ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโล-

จิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า วิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 1000 : 250$ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซล ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 1000 : 500$ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด รองลงมาได้แก่วิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว ปรากฏว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีซิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ขนาดกลุ่มตัวอย่าง 6 ระดับ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ในขณะที่อัตราความคลาดเคลื่อนประเภท 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

ผู้วิจัยนำค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยที่ศึกษา 4 ตัว ไปเปรียบเทียบกับเกณฑ์ที่ระดับ 10% ผลการเปรียบเทียบแสดงดังตารางที่ 60

ตารางที่ 60 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ 4 วิธีกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เงื่อนไขของปัจจัยที่ศึกษา

ปัจจัย	วิธีการตรวจสอบ DIF				เกณฑ์ของ E_1(%)
	วิธี MSIB	วิธี SIB	วิธี MH	วิธี LR	
	E_1 (%)	E_1 (%)	E_1 (%)	E_1 (%)	
TI					
a_L , b_L	9.50	2.94	4.53	9.94	10
a_L , b_M	10.64	5.00	6.31	9.25	10
a_L , b_H	10.78	4.28	7.39	9.33	10
a_M , b_L	9.28	4.39	4.92	8.06	10
a_M , b_M	7.36	4.17	3.69	6.19	10
a_M , b_H	8.64	5.64	5.56	8.17	10
a_H , b_L	7.92	3.89	2.31	6.36	10
a_H , b_M	7.56	3.58	3.75	5.36	10
a_H , b_H	8.25	4.89	5.64	7.92	10
TL					
30 ข้อ	9.62	4.84	5.07	8.04	10
60 ข้อ	8.14	3.78	4.73	7.64	10
%DIF					
5%	9.02	4.35	5.06	7.61	10
10%	7.87	4.64	5.19	7.63	10
20%	9.75	3.94	4.44	8.29	10
N_R : N_F					
250 : 250	8.69	4.69	5.52	9.28	10
500 : 250	10.00	4.61	4.69	8.09	10
500 : 500	9.04	4.56	4.67	8.63	10
1000 : 250	10.04	5.02	4.74	7.94	10
1000 : 500	8.24	3.59	5.61	6.93	10
1000 : 1000	7.28	3.39	4.17	6.19	10
ค่าเฉลี่ยรวม	8.88	4.31	4.90	7.84	10

จากตารางที่ 60 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธีกับเกณฑ์ที่ระดับ 10% ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

ภายใต้เงื่อนไขของลักษณะของข้อสอบ (TI) อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ในช่วงประมาณ 7.36% ถึง 10.78% ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าอยู่ในช่วงประมาณ 2.94% ถึง 5.64% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าอยู่ในช่วงประมาณ 2.31% ถึง 7.39% ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าอยู่ในช่วงประมาณ 5.36% ถึง 9.94% จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% เกือบทุกเงื่อนไขของลักษณะของข้อสอบ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของอีก 3 วิธีมีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% ในทุกเงื่อนไขของลักษณะของข้อสอบ

ภายใต้เงื่อนไขของความยาวของแบบสอบ (TL) อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ในช่วงประมาณ 8.14% ถึง 9.62% ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าอยู่ในช่วงประมาณ 3.78% ถึง 4.84% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าอยู่ในช่วงประมาณ 4.73% ถึง 5.07% ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าอยู่ในช่วงประมาณ 7.64% ถึง 8.04% จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% ในทุกเงื่อนไขของความยาวของแบบสอบ

ภายใต้เงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ในแบบสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ในช่วงประมาณ 7.87% ถึง 9.75% ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าอยู่ในช่วงประมาณ 3.94% ถึง 4.64% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าอยู่ในช่วงประมาณ 4.44% ถึง 5.19% ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าอยู่ในช่วงประมาณ 7.61% ถึง 8.29% จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% ในทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่าง (N_R : N_F) อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ในช่วงประมาณ 7.28% ถึง 10.04% ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าอยู่ในช่วงประมาณ 3.39% ถึง 5.02% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าอยู่ในช่วงประมาณ 4.17% ถึง 5.61% ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าอยู่ในช่วงประมาณ

6.19% ถึง 9.28% จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของอีก 3 วิธีมีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% ในทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

จากผลดังกล่าว เมื่อพิจารณาในทุกเงื่อนไขของปัจจัยที่ศึกษา ปรากฏว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10%