บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิบเทสท์ปรับใหม่ วิธีซิบเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยศึกษาภายใต้การจำลองข้อมูลด้วยการจัดกระทำปัจจัยที่แปรเปลี่ยน 4 ตัว คือ ลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ในแต่ละปัจจัยยังแบ่งออกเป็นระดับย่อย ๆ ดังนี้ 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) ความยาวของแบบสอบ 2 ระดับ (30 ข้อ และ 60 ข้อ) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ (5%, 10% และ 20%) และขนาดกลุ่มตัวอย่าง 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)

ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลจำลอง โดยใช้โปรแกรม IRTDATA (Johanson, 1992) จำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ (3PLM-c) แล้วนำข้อมูลที่จำลองไปจัดกระทำตามปัจจัยที่ศึกษา 4 ตัว ในการจัดกระทำข้อมูลดังกล่าว ผู้วิจัยใช้โปรแกรม BILOG version 3.04 (Mislevy and Bock, 1990) ประมาณค่าพารามิเตอร์ของข้อสอบ โดยเลือกโมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ แล้วปรับเทียบสเกลพารามิเตอร์ของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบให้อยู่บนเมทริกซ์เดียวกันด้วยวิธีโค้งลักษณะแบบสอบ (test characteristic curve; TCC) ของ Stocking และ Lord (1983 cited in Kim and Cohen, 1992a) ภายใต้โมเดลโลจิสติกแบบ 2 พารามิเตอร์ โดยใช้โปรแกรม EQUATE version 2.0 (Baker, 1993) ต่อจากนั้นจึงตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ($\hat{a}_{iR} \neq \hat{a}_{iF}$) ด้วยวิธีการวัดพื้นที่ซนิดไม่คิดเครื่องหมายของ Raju (1990) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ (3PLM-c) โดยใช้โปรแกรม IRTDIF version 1.0 (Kim and Cohen, 1992b) แล้วจึงนำผลการตรวจสอบดังกล่าวไปจัดกระทำข้อมูลตามปัจจัยที่ศึกษา 4 ตัว โดยใช้โปรแกรม SPSS for windows version 7.52 สุ่มข้อมูล รวมข้อมูลที่จัดกระทำทั้งหมด 324 เงื่อนไข ($9 \times 2 \times 3 \times 6$)

ผู้วิจัยนำข้อมูลที่จัดกระทำ 324 เงื่อนไข ไปคำนวณการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป 4 วิธี คือ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยสองวิธีแรกใช้โปรแกรม SIBTEST ของ Stout และ Roussos version 1.1 (1992) วิธีต่อมาใช้โปรแกรม MHDIF version 1.0 (Fidalgo, 1995) และวิธีสุดท้ายใช้โปรแกรม SPSS/PC+ version 4.01

ในการวิเคราะห์ค่าอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีดังกล่าว ผู้วิจัยใช้วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมายของ Raju เป็นเกณฑ์ในการเปรียบเทียบ ส่วนการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธีภายใต้ปัจจัยที่ศึกษา 4 ตัว ผู้วิจัยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติจะใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว (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

สรุปผลการวิจัย

ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยศึกษาภายใต้การจัดกระทำปัจจัย 4 ตัว คือ ลักษณะของข้อสอบ 9 ลักษณะ ความยาวของแบบสอบ 2 ระดับ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ และขนาดกลุ่มตัวอย่าง 6 ระดับ สรุปรวมได้ดังตารางที่ 61, 62, 63 และ 64

ตารางที่ 61 สรุปผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบ (p) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้ทุกเงื่อนไขของปัจจัยที่ศึกษา

ปัจจัย	วิธีการตรวจสอบ DIF							
	วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
	P	E_1	P	E_1	P	E_1	P	E_1
TI								
a_L , b_L	26.36	9.50	18.03	2.94	19.17	4.53	24.33	9.94
a_L , b_M	53.22	10.64	12.69	5.00	19.22	6.31	39.75	9.25
a_L , b_H	27.53	10.78	22.89	4.28	18.53	7.39	22.39	9.33
a_M , b_L	25.11	9.28	6.72	4.39	5.78	4.92	15.54	8.06
a_M , b_M	45.81	7.36	25.17	4.17	29.14	3.69	49.28	6.19
a_M , b_H	10.86	8.64	5.11	5.64	5.53	5.56	30.69	8.17
a_H , b_L	8.08	7.92	1.83	3.89	15.97	2.31	19.48	6.36
a_H , b_M	37.64	7.56	16.61	3.58	26.08	3.75	49.75	5.36
a_H , b_H	23.17	8.25	10.92	4.89	8.61	5.64	19.69	7.92
TL								
30 ข้อ	28.17	9.62	13.05	4.84	16.77	5.07	31.38	8.04
60 ข้อ	29.12	8.14	13.61	3.78	16.12	4.73	28.82	7.64
%DIF								
5%	27.29	9.02	14.94	4.35	20.94	5.06	38.24	7.61
10%	30.08	7.87	14.19	4.64	17.13	5.19	31.00	7.63
20%	28.56	9.75	10.85	3.94	11.28	4.44	21.06	8.29
N_R : N_F								
250 : 250	20.00	8.69	9.09	4.69	10.80	5.52	24.38	9.28
500 : 250	21.09	10.00	10.17	4.61	12.72	4.69	22.63	8.09
500 : 500	30.41	9.04	12.63	4.56	13.09	4.67	27.93	8.63
1000 : 250	24.28	10.04	12.65	5.02	13.41	4.74	24.37	7.94
1000 : 500	33.94	8.24	14.17	3.59	22.54	5.61	34.54	6.93
1000 : 1000	42.13	7.28	21.28	3.39	26.13	4.17	46.76	6.19
ค่าเฉลี่ยรวม	28.64	8.88	13.33	4.31	16.45	4.90	30.10	7.84

ตารางที่ 62 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (p) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของแต่ละวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน

ปัจจัย		วิธีการตรวจสอบ DIF							
		วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
		P	E_1	P	E_1	P	E_1	P	E_1
❑ a_L , b_L	▲ a_L , b_M	–	0	0	0	0	0	0	0
	▲ a_L , b_H	0	0	0	0	0	–	0	0
	▲ a_M , b_L	0	0	0	0	0	0	0	0
	▲ a_M , b_M	–	0	0	0	0	0	–	+
	▲ a_M , b_H	0	0	+	–	0	0	0	0
	▲ a_H , b_L	0	0	+	0	0	0	0	+
	▲ a_H , b_M	0	0	0	0	0	0	–	+
	▲ a_H , b_H	0	0	0	0	0	0	0	0
❑ a_L , b_M	▲ a_L , b_L	–	0	0	0	0	0	0	0
	▲ a_L , b_H	–	0	0	0	0	0	0	0
	▲ a_M , b_L	–	0	0	0	0	0	+	0
	▲ a_M , b_M	0	0	–	0	0	0	0	0
	▲ a_M , b_H	–	0	+	0	0	0	0	0
	▲ a_H , b_L	–	0	+	0	0	+	+	0
	▲ a_H , b_M	0	0	0	0	0	0	0	+
	▲ a_H , b_H	–	0	0	0	0	0	+	0
❑ a_L , b_H	▲ a_L , b_L	0	0	0	0	0	+	0	0
	▲ a_L , b_M	–	0	0	0	0	0	0	0
	▲ a_M , b_L	0	0	+	0	0	0	0	0
	▲ a_M , b_M	0	0	0	0	0	+	–	+
	▲ a_M , b_H	0	0	+	0	0	0	0	0
	▲ a_H , b_L	–	0	+	0	0	+	0	0
	▲ a_H , b_M	0	0	0	0	0	+	–	+
	▲ a_H , b_H	0	0	+	0	0	0	0	0

ตารางที่ 62 (ต่อ)

ปัจจัย		วิธีการตรวจสอบ DIF							
		วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
		P	E_1	P	E_1	P	E_1	P	E_1
❑ a_M , b_L	▲ a_L , b_L	0	0	0	0	0	0	0	0
	▲ a_L , b_M	–	0	0	0	0	0	–	0
	▲ a_L , b_H	0	0	–	0	0	0	0	0
	▲ a_M , b_M	–	0	–	0	–	0	–	0
	▲ a_M , b_H	0	0	0	0	0	0	0	0
	▲ a_H , b_L	0	0	0	0	0	0	0	0
	▲ a_H , b_M	0	0	0	0	–	0	–	0
	▲ a_H , b_H	0	0	0	0	0	0	0	0
❑ a_M , b_M	▲ a_L , b_L	–	0	0	0	0	0	+	–
	▲ a_L , b_M	0	0	+	0	0	0	0	0
	▲ a_L , b_H	0	0	0	0	0	–	+	–
	▲ a_M , b_L	–	0	+	0	+	0	+	0
	▲ a_M , b_H	–	0	+	0	+	0	0	0
	▲ a_H , b_L	–	0	+	0	0	0	+	0
	▲ a_H , b_M	0	0	0	0	0	0	0	0
	▲ a_H , b_H	–	0	+	0	+	0	+	0
❑ a_M , b_H	▲ a_L , b_L	0	0	–	+	0	0	0	0
	▲ a_L , b_M	–	0	0	0	0	0	0	0
	▲ a_L , b_H	0	0	–	0	0	0	0	0
	▲ a_M , b_L	0	0	0	0	0	0	0	0
	▲ a_M , b_M	–	0	–	0	–	0	0	0
	▲ a_H , b_L	0	0	0	0	0	+	0	0
	▲ a_H , b_M	–	0	0	0	–	0	0	0
	▲ a_H , b_H	0	0	0	0	0	0	0	0

ตารางที่ 62 (ต่อ)

ปัจจัย		วิธีการตรวจสอบ DIF							
		วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
		P	E_1	P	E_1	P	E_1	P	E_1
❑ a_H , b_L	▲ a_L , b_L	0	0	–	0	0	0	0	–
	▲ a_L , b_M	–	0	0	0	–	–	–	0
	▲ a_L , b_H	–	0	–	0	–	–	0	0
	▲ a_M , b_L	0	0	0	0	0	0	0	0
	▲ a_M , b_M	–	0	–	0	0	0	–	0
	▲ a_M , b_H	0	0	0	0	–	–	0	0
	▲ a_H , b_M	–	0	–	0	0	0	–	0
	▲ a_H , b_H	0	0	0	0	–	–	0	0
❑ a_H , b_M	▲ a_L , b_L	0	0	0	0	0	0	+	–
	▲ a_L , b_M	0	0	0	0	0	0	0	–
	▲ a_L , b_H	0	0	0	0	0	–	+	–
	▲ a_M , b_L	0	0	0	0	+	0	+	0
	▲ a_M , b_M	0	0	0	0	0	0	0	0
	▲ a_M , b_H	+	0	0	0	+	0	0	0
	▲ a_H , b_L	+	0	+	0	0	0	+	0
	▲ a_H , b_H	0	0	0	0	+	0	+	0
❑ a_H , b_H	▲ a_L , b_L	0	0	0	0	0	0	0	0
	▲ a_L , b_M	–	0	0	0	0	0	0	0
	▲ a_L , b_H	0	0	–	0	0	0	0	0
	▲ a_M , b_L	0	0	0	0	0	0	0	0
	▲ a_M , b_M	–	0	–	0	–	0	–	0
	▲ a_M , b_H	0	0	0	0	0	0	0	0
	▲ a_H , b_L	0	0	0	0	0	+	0	0
	▲ a_H , b_M	0	0	0	0	–	0	–	0

ตารางที่ 62 (ต่อ)

ปัจจัย		วิธีการตรวจสอบ DIF							
		วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
		P	E_1	P	E_1	P	E_1	P	E_1
❑ 30 ข้อ	▲ 60 ข้อ	0	+	0	+	0	0	0	0
❑ 5%	▲ 10%	0	0	0	0	0	0	0	0
	▲ 20%	0	0	0	0	+	0	+	0
❑ 10%	▲ 5%	0	0	0	0	0	0	0	0
	▲ 20%	0	–	0	0	0	0	+	0
❑ 20%	▲ 5%	0	0	0	0	–	0	–	0
	▲ 10%	0	+	0	0	0	0	–	0
❑ 250 : 250	▲ 500 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 250	0	0	0	0	0	0	0	0
	▲ 1000 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 1000	–	0	–	0	–	0	–	+
❑ 500 : 250	▲ 250 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 250	0	0	0	0	0	0	0	0
	▲ 1000 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 1000	–	0	–	0	–	0	–	0
❑ 500 : 500	▲ 250 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 250	0	0	0	0	0	0	0	0
	▲ 1000 : 250	0	0	0	0	0	0	0	0
	▲ 1000 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 1000	0	0	0	0	–	0	–	–
❑ 1000 : 250	▲ 250 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 1000	–	0	0	0	–	0	–	0

ตารางที่ 62 (ต่อ)

ปัจจัย		วิธีการตรวจสอบ DIF							
		วิธี MSIB		วิธี SIB		วิธี MH		วิธี LR	
		P	E_1	P	E_1	P	E_1	P	E_1
❑ 1000 : 500	▲ 250 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 250	0	0	0	0	0	0	0	0
	▲ 500 : 500	0	0	0	0	0	0	0	0
	▲ 1000 : 250	0	0	0	0	0	0	0	0
	▲ 1000 : 1000	0	0	0	0	0	0	0	0
❑ 1000 : 1000	▲ 250 : 250	+	0	+	0	+	0	+	–
	▲ 500 : 250	+	0	+	0	+	0	+	0
	▲ 500 : 500	0	0	0	0	+	0	+	–
	▲ 1000 : 250	+	0	0	0	+	0	+	0
	▲ 1000 : 500	0	0	0	0	0	0	0	0

0 หมายถึง P (หรือ E_1) ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ❑ มีค่าเท่ากับ P (หรือ E_1) ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ▲

\+ หมายถึง P (หรือ E_1) ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ❑ มีค่าสูงกว่า P (หรือ E_1) ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ▲

– หมายถึง P (หรือ E_1) ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ❑ มีค่าต่ำกว่า P (หรือ E_1) ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ▲

ตารางที่ 63 สรุปผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน

ปัจจัย	วิธีการตรวจสอบ DIF											
	■วิธี MSIB			■ วิธี SIB			■ วิธี MH			■ วิธี LR		
	● วิธี SIB	● วิธี MH	● วิธี LR	● วิธี MSIB	● วิธี MH	● วิธี LR	● วิธี MSIB	● วิธี SIB	● วิธี LR	● วิธี MSIB	● วิธี SIB	● วิธี MH
a_L , b_L	0	0	0	0	0	0	0	0	0	0	0	0
a_L , b_M	+	+	0	–	0	–	–	0	–	0	+	+
a_L , b_H	0	0	0	0	0	0	0	0	0	0	0	0
a_M , b_L	+	+	0	–	0	0	–	0	0	0	0	0
a_M , b_M	+	0	0	–	0	–	0	0	–	0	+	+
a_M , b_H	0	0	–	0	0	–	0	0	–	+	+	+
a_H , b_L	0	0	–	0	–	–	0	+	0	+	+	0
a_H , b_M	+	0	0	–	0	–	0	0	–	0	+	+
a_H , b_H	+	+	0	–	0	0	–	0	–	0	0	+
30 ข้อ	+	+	0	–	0	–	–	0	–	0	+	+
60 ข้อ	+	+	0	–	0	–	–	0	–	0	+	+
5%	+	0	0	–	0	–	0	0	–	+	+	+
10%	+	+	0	–	0	–	–	0	–	0	+	+
20%	+	+	+	–	0	–	–	0	–	–	+	+
250 : 250	+	+	0	–	0	–	–	0	–	0	+	+
500 : 250	+	0	0	–	0	–	0	0	–	0	+	+
500 : 500	+	+	0	–	0	–	–	0	–	0	+	+
1000 : 250	+	+	0	–	0	–	–	0	–	0	+	+
1000 : 500	+	0	0	–	0	0	0	0	0	0	0	0
1000 :1000	+	0	0	–	0	–	0	0	–	0	+	+

0 หมายถึง P ของวิธีการตรวจสอบ ■ มีค่าเท่ากับ P ของวิธีการตรวจสอบ ●

+ หมายถึง P ของวิธีการตรวจสอบ ■ มีค่าสูงกว่า P ของวิธีการตรวจสอบ ●

– หมายถึง P ของวิธีการตรวจสอบ ■ มีค่าต่ำกว่า P ของวิธีการตรวจสอบ ●

ตารางที่ 64 สรุปผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน

ปัจจัย	วิธีการตรวจสอบ DIF											
	■วิธี MSIB			■ วิธี SIB			■ วิธี MH			■ วิธี LR		
	● วิธี SIB	● วิธี MH	● วิธี LR	● วิธี MSIB	● วิธี MH	● วิธี LR	● วิธี MSIB	● วิธี SIB	● วิธี LR	● วิธี MSIB	● วิธี SIB	● วิธี MH
a_L , b_L	+	+	0	–	0	–	–	0	–	0	+	+
a_L , b_M	+	+	0	–	0	–	–	0	–	0	+	+
a_L , b_H	+	+	0	+	–	–	–	+	0	0	+	0
a_M , b_L	+	+	0	–	0	–	–	0	–	0	+	+
a_M , b_M	+	+	0	–	0	0	–	0	–	0	0	+
a_M , b_H	+	+	0	–	0	0	–	0	0	0	0	0
a_H , b_L	+	+	0	–	0	–	–	0	–	0	+	+
a_H , b_M	+	+	0	–	0	0	–	0	0	0	0	0
a_H , b_H	+	+	0	–	0	0	–	0	0	0	0	0
30 ข้อ	+	+	+	–	0	–	–	0	–	–	+	+
60 ข้อ	+	+	0	–	0	–	–	0	–	0	+	+
5%	+	+	0	–	0	–	–	0	–	0	+	+
10%	+	+	0	–	0	–	–	0	–	0	+	+
20%	+	+	0	–	0	–	–	0	–	0	+	+
250 : 250	+	+	0	–	0	–	–	0	–	0	+	+
500 : 250	+	+	0	–	0	–	–	0	–	0	+	+
500 : 500	+	+	0	–	0	–	–	0	–	0	+	+
1000 : 250	+	+	0	–	0	–	–	0	–	0	+	+
1000 : 500	+	+	0	–	–	–	–	+	0	0	+	0
1000 :1000	+	+	0	–	0	–	–	0	–	0	+	+

0 หมายถึง E_1 ของวิธีการตรวจสอบ ■ มีค่าเท่ากับ E_1 ของวิธีการตรวจสอบ ●

\+ หมายถึง E_1 ของวิธีการตรวจสอบ ■ มีค่าสูงกว่า E_1 ของวิธีการตรวจสอบ ●

– หมายถึง E_1 ของวิธีการตรวจสอบ ■ มีค่าต่ำกว่า E_1 ของวิธีการตรวจสอบ ●

ผลการวิจัยสรุปเป็น 2 ประเด็นใหญ่ ๆ ดังนี้

1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

1.1 ปัจจัยทางด้านลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อลักษณะของข้อสอบ a ต่ำกับ b ปานกลางจะส่งผลให้อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด เมื่อลักษณะของข้อสอบ a ปานกลางกับ b ปานกลางจะส่งผลให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าสูงสุด เมื่อลักษณะของข้อสอบ a สูงกับ b ปานกลางจะส่งผลให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงสุด เมื่อพิจารณาในทุกเงื่อนไขของลักษณะของข้อสอบ ปรากฏว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นในทุกเงื่อนไขของลักษณะของข้อสอบที่มีค่าความยากปานกลาง (a_L, b_M; a_M, b_M และ a_H, b_M)

1.2 ปัจจัยทางด้านความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

1.3 ปัจจัยทางด้านสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ แต่จะมีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนลดลงจาก 20% เป็น 5% จะส่งผลให้อำนาจการทดสอบของ 2 วิธีดังกล่าวมีค่าเพิ่มขึ้นประมาณ 10% และ 17% ตามลำดับ

1.4 ปัจจัยทางด้านขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นจาก $N_R : N_F = 250 : 250$ เป็น $N_R : N_F = 1000 : 1000$ จะส่งผลให้อำนาจการทดสอบของทั้ง 4 วิธีดังกล่าวมีค่าเพิ่มขึ้นสูงสุด โดยเพิ่มขึ้นประมาณ 22%, 12%, 15% และ 22% ตามลำดับ

1.5 ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ปรากฏว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกันภายใต้เกือบ

ทุกเงื่อนไข ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ภายใต้เกือบทุกเงื่อนไข

2. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

2.1 ปัจจัยทางด้านลักษณะของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อลักษณะของข้อสอบ a ปานกลางกับ b สูงจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงสุด เมื่อลักษณะของข้อสอบ a ต่ำกับ b สูงจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1ของวิธีแมนเทล-แฮนส์เซลมีค่าสูงสุด เมื่อลักษณะของข้อสอบ a ต่ำกับ b ต่ำจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงสุด เมื่อพิจารณาในทุกเงื่อนไขของลักษณะของข้อสอบ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นในทุกเงื่อนไขของลักษณะของข้อสอบที่มีค่าอำนาจจำแนกต่ำ (a_L, b_L; a_L, b_M และ a_L, b_H)

2.2 ปัจจัยทางด้านความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก แต่จะมีผลต่อวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบจาก 30 ข้อ เป็น 60 ข้อ จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของ 2 วิธีดังกล่าวมีค่าลดลงต่ำสุด โดยลดลงประมาณ 1% เท่ากันทั้ง 2 วิธี

2.3 ปัจจัยทางด้านสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก แต่จะมีผลต่อวิธีซิปเทสท์ปรับใหม่ กล่าวคือ เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนเพิ่มขึ้นจาก 10% เป็น 20% จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีดังกล่าวมีค่าเพิ่มขึ้นสูงสุด โดยเพิ่มขึ้นประมาณ 2%

2.4 ปัจจัยทางด้านขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซล แต่จะมีผลต่อวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นจาก $N_R : N_F = 250 : 250$ เป็น $N_R : N_F =$

1000 : 1000 จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีดังกล่าวมีค่าลดลงต่ำสุด โดยลดลงประมาณ 3%

2.5 ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันและขนาดกลุ่มตัวอย่าง ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ภายใต้เกือบทุกเงื่อนไข

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธีภายใต้เงื่อนไขของปัจจัยที่ศึกษา 324 เงื่อนไขกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าอยู่ภายในเกณฑ์ที่กำหนดดังกล่าว

เมื่อพิจารณาภาพรวมของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง พบว่า วิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีความถูกต้องแม่นยำในการตรวจสอบได้เท่าเทียมกัน และทั้ง 2 วิธีมีความถูกต้องแม่นยำในการตรวจสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล

รายละเอียดของผลการวิจัยสรุปได้ดังนี้

1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

1.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงสุด (ประมาณร้อยละ 53.22) 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 (ประมาณร้อยละ 45.81, 37.64, 27.53, 26.36, 25.11, 23.17, 10.86 และ 8.08 ตามลำดับ) สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงสุด (ประมาณร้อยละ 25.17) รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ 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 (ประมาณร้อยละ 22.89, 18.03, 16.61, 12.69, 10.92, 6.72, 5.11 และ 1.83 ตามลำดับ) ส่วนวิธีแมนเทล-แฮนส์เซล พบว่า อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_M มีค่าสูงสุด (ประมาณร้อยละ 29.14) 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 (ประมาณร้อยละ 26.08, 19.22, 19.17, 18.53, 15.97, 8.61, 5.78 และ 5.53 ตามลำดับ) สำหรับวิธีการถดถอยโลจิสติก พบว่า อำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงสุด (ประมาณร้อยละ 49.75) รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_M, b_M; a_L, 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 (ประมาณร้อยละ 49.28, 39.75, 30.69, 24.33, 22.39, 19.69, 19.48 และ 15.54 ตามลำดับ) เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีดังกล่าวภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและลักษณะของข้อสอบมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ (1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ (2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน 9 ลักษณะ และ (3) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ปรากฏว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 4 วิธี ดังนั้นจึงนำอำนาจการทดสอบของแต่ละวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ 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 สำหรับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ 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 สำหรับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H และ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบ 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 ส่วนอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_M, b_H; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_H และ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบ a_M, b_M และ a_H, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_M, b_H และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบ a_M, b_M และ 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 สำหรับอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ลักษณะของข้อสอบ a_M, b_L; a_H, b_L และ a_H, b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ปรากฏว่า ภายใต้ลักษณะของข้อสอบ a_L, b_L และ a_L, b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ 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 ดังนั้นจึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบดังกล่าวมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

ภายใต้ลักษณะของข้อสอบ a_L, b_M อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_L อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_M อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-

แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_H อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_L อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_M อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_H อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

กรณีที่ 3 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a_L, b_M มีค่าสูงสุด โดยมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ ส่วนอำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำสุด โดยมีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ

1.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 28.17 และ 29.12 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 13.05 และ 13.61 ตามลำดับ สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 16.77 และ 16.12 ตามลำดับ สำหรับวิธีการถดถอยโลจิสติก พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 31.38 และ 28.82 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีดังกล่าวภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ปรากฏว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออำนาจการทดสอบ เมื่อพิจารณาผลของความยาวของแบบสอบ พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบ และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออำนาจการทดสอบ ดังนั้นจึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน 2 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ผลปรากฏดังนี้

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน 2 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ มาเปรียบเทียบเป็นรายคู่ ปรากฏว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ ได้ผลสอดคล้องกัน กล่าวคือ อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน และทั้ง 2 วิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน

1.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าสูงสุด (ประมาณร้อยละ 30.08)

รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% และ 5% ตามลำดับ (ประมาณร้อยละ 28.56 และ 27.29 ตามลำดับ) ส่วนวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบ ที่ทำหน้าที่ต่างกัน 5% มีค่าสูงสุด (ประมาณร้อยละ 14.94) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ตามลำดับ (ประมาณร้อยละ 14.19 และ 10.85 ตามลำดับ) สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงสุด (ประมาณร้อยละ 20.94) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ตามลำดับ (ประมาณร้อยละ 17.13 และ 11.28 ตามลำดับ) ส่วนวิธีการถดถอยโลจิสติก พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงสุด (ประมาณร้อยละ 38.24) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ตามลำดับ (ประมาณร้อยละ 31.00 และ 21.06 ตามลำดับ) เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ (1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ (2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ และ (3) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์มีค่าไม่แตกต่างกัน แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของ 2 วิธีดังกล่าว ส่วนอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีผลต่ออำนาจการทดสอบของทั้ง 2 วิธีดังกล่าว ดังนั้นจึงนำอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงกว่าอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าต่ำกว่าอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% มีค่าไม่แตกต่างกัน

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ปรากฏว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% มาเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% ได้ผลสอดคล้องกัน กล่าวคือ อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน แต่ทั้ง 2 วิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธี แมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน

กรณีที่ 3 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey ผลปรากฏว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% มีค่าสูงสุด โดยมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ ในขณะเดียวกัน อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าต่ำสุด โดยมีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ

1.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ ปรากฏว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 1000 : 1000$ มีค่าสูงสุด (ประมาณร้อยละ 42.13) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 1000 : 500$, $500 : 500$, $1000 : 250$, $500 : 250$ และ $250 : 250$

ตามลำดับ (ประมาณร้อยละ 33.94, 30.41, 24.28, 21.09 และ 20.00 ตามลำดับ) สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด (ประมาณร้อยละ 21.28) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 1000 : 250, 500 : 500, 500 : 250 และ 250 : 250 ตามลำดับ (ประมาณร้อยละ 14.17, 12.65, 12.63, 10.17 และ 9.09 ตามลำดับ) สำหรับวิธีแมนเทล-แฮนส์เซล ปรากฏว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด (ประมาณร้อยละ 26.13) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 1000 : 250, 500 : 500, 500 : 250 และ 250 : 250 ตามลำดับ (ประมาณร้อยละ 22.54, 13.41, 13.09, 12.72 และ 10.80 ตามลำดับ) สำหรับวิธีการถดถอยโลจิสติก พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าสูงสุด (ประมาณร้อยละ 46.76) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500, 500 : 500, 250 : 250, 1000 : 250 และ 500 : 250 ตามลำดับ (ประมาณร้อยละ 34.54, 27.93, 24.38, 24.37 และ 22.63 ตามลำดับ) เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ปรากฏว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและขนาดกลุ่มตัวอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและขนาดกลุ่มตัวอย่างไม่มีผลร่วมกันต่ออำนาจการทดสอบ เมื่อพิจารณาผลของขนาดกลุ่มตัวอย่าง ปรากฏว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบ และเมื่อพิจารณาผลของวิธีการตรวจสอบ ปรากฏว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออำนาจการทดสอบ ดังนั้นจึงเปรียบเทียบ 2 กรณี คือ (1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ และ (2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 4 วิธี ดังนั้นจึงนำอำนาจการทดสอบของแต่ละวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ มาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250, 500 : 250 และ 1000 : 250 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 1000 : 1000 มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250 และ 500 : 250 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ขนาดกลุ่มตัวอย่าง 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

อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 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

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ปรากฏว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างทั้ง 6 ระดับ มาเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

ภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250, 500 : 500, 1000 : 250 และ 1000 : 1000 ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน แต่ทั้ง 2 วิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีอำนาจการทดสอบไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 500 : 250$ อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่มีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F = 1000 : 500$ อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

2. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงสุด (ประมาณร้อยละ 10.78) รองลงมาได้แก่ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_L, b_L; a_M, b_L; a_M, b_H; a_H, b_H; a_H, b_L; a_H, b_M และ a_M, b_M ตามลำดับ (ประมาณร้อยละ 10.64, 9.50, 9.28, 8.64, 8.25, 7.92, 7.56 และ 7.36 ตามลำดับ) สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าสูงสุด (ประมาณร้อยละ 5.64) 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 (ประมาณร้อยละ 5.00, 4.89, 4.39, 4.28, 4.17, 3.89, 3.58 และ 2.94 ตามลำดับ) สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงสุด (ประมาณร้อยละ 7.39) 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 (ประมาณร้อยละ 6.31, 5.64, 5.56, 4.92, 4.53, 3.75, 3.69 และ 2.31 ตามลำดับ) ส่วนวิธีการถดถอยโลจิสติก พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L มีค่าสูงสุด (ประมาณร้อยละ 9.94) 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 (ประมาณร้อยละ 9.33, 9.25, 8.17, 8.06, 7.92, 6.36, 6.19 และ 5.36 ตามลำดับ) เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีดังกล่าวภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่าผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและลักษณะของข้อสอบมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ (1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ (2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน 9 ลักษณะ และ (3) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ลักษณะของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธี ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า วิธีซิปเทสท์ปรับใหม่ไม่พบความแตกต่าง ส่วนวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกได้ผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ภายใต้ลักษณะของข้อสอบ a_M, b_H มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบ 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 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_L, b_M; a_M, b_H และ a_H, b_H มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบ a_H, b_L อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ลักษณะของข้อสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

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

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบเดียวกัน 9 ลักษณะ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า ภายใต้ลักษณะของข้อสอบทั้ง 9 ลักษณะ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบ 9 ลักษณะมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏดังนี้

ภายใต้ลักษณะของข้อสอบ a_L, b_L; a_L, b_M; a_M, b_L และ a_H, b_L ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_L, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซลมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_M อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_M, b_H และ a_H, b_M ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

ภายใต้ลักษณะของข้อสอบ a_H, b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

กรณีที่ 3 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey ผลปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ลักษณะของข้อสอบ a_L, b_H มีค่าสูงสุด โดยมีค่าสูงกว่าอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลเกือบทุกเงื่อนไขของการตรวจสอบ ในขณะเดียวกัน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้ลักษณะของข้อสอบ a_H, b_L มีค่าต่ำสุด โดยมีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ความยาวของแบบสอบ 30 ข้อ มีค่าประมาณร้อยละ 9.62 และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้อ มีค่าประมาณร้อยละ 8.14 สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 30 ข้อ มีค่าประมาณร้อยละ 4.84 และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้อ มีค่าประมาณร้อยละ 3.78 สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1

ภายใต้ความยาวของแบบสอบ 30 ข้อ มีค่าประมาณร้อยละ 5.07 และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้อ มีค่าประมาณร้อยละ 4.73 สำหรับวิธีการถดถอยโลจิสติก พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 30 ข้อ มีค่าประมาณร้อยละ 8.04 และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้อ มีค่าประมาณร้อยละ 7.64 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีดังกล่าวภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ปรากฏว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของความยาวของแบบสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ความยาวของแบบสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้นจึงเปรียบเทียบ 2 กรณี คือ (1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ และ (2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน 2 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ โดยใช้สถิติ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน ผลปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของ 2 วิธีดังกล่าว ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ภายใต้ความยาวของแบบสอบ 30 ข้อมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 60 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ปัจจัยทางด้านความยาวของแบบสอบมีผลต่ออัตราความคลาดเคลื่อนประเภท ที่ 1 ของ 2 วิธีดังกล่าว

กรณีที่ 2 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบระดับเดียวกัน 2 ระดับโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า มีค่า

แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ มาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

ภายใต้ความยาวของแบบสอบ 30 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน

ภายใต้ความยาวของแบบสอบ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน แต่ทั้ง 2 วิธีดังกล่าวมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน

2.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนต่างกัน 3 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าสูงสุด (ประมาณร้อยละ 9.75) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 10% ตามลำดับ (ประมาณร้อยละ 9.02 และ 7.87 ตามลำดับ) สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าสูงสุด (ประมาณร้อยละ 4.64) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 20% ตามลำดับ (ประมาณร้อยละ 4.35 และ 3.94 ตามลำดับ) ส่วนวิธีแมนเทล-แฮนส์เซล พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% มีค่าสูงสุด (ประมาณร้อยละ 5.19) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% และ 20% ตามลำดับ (ประมาณร้อยละ 5.06 และ 4.44 ตามลำดับ) สำหรับวิธีการถดถอยโลจิสติก พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าสูงสุด (ประมาณร้อยละ 8.29) รองลงมาได้แก่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 5% ตามลำดับ (ประมาณร้อยละ 7.63 และ 7.61 ตามลำดับ) เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวน

ต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ (1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ (2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับ และ (3) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของ 3 วิธีดังกล่าว ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสท์ปรับใหม่มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีดังกล่าว ดังนั้นจึงเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

กรณีที่ 2 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันระดับเดียวกัน 3 ระดับโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ปรากฏว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% มาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสท์-

ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน แต่ทั้ง 2 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน

กรณีที่ 3 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบเป็นรายคู่ของ Tukey ผลปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าสูงสุด โดยมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ ในขณะเดียวกัน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% มีค่าต่ำสุด โดยมีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกในทุกเงื่อนไขของการตรวจสอบ

2.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 250 มีค่าสูงสุด (ประมาณร้อยละ 10.04) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 250, 500 : 500, 250 : 250, 1000 : 500 และ 1000 : 1000 ตามลำดับ (ประมาณร้อยละ 10.00, 9.04, 8.69, 8.24 และ 7.28 ตามลำดับ) สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 250 มีค่าสูงสุด (ประมาณร้อยละ 5.02) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 500 และ 1000 : 1000 ตามลำดับ (ประมาณร้อยละ 4.69, 4.61, 4.56, 3.59 และ 3.39 ตามลำดับ) สำหรับวิธีแมนเทล-แฮนส์เซล พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500 มีค่าสูงสุด (ประมาณร้อยละ 5.61) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 1000 : 250, 500 : 250, 500 : 500 และ 1000 : 1000 (ประมาณร้อยละ 5.52, 4.74, 4.69, 4.67 และ 4.17 ตามลำดับ) สำหรับวิธีการถดถอยโลจิสติก พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 มีค่าสูงสุด (ประมาณร้อยละ 9.28) รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 500, 500 : 250, 1000 : 250, 1000 : 500 และ 1000 : 1000 (ประมาณร้อยละ 8.63, 8.09, 7.94, 6.93 และ 6.19 ตามลำดับ) เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่าง

ต่างกัน 6 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ปรากฏว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและขนาดกลุ่มตัวอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและขนาดกลุ่มตัวอย่างไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของขนาดกลุ่มตัวอย่าง ปรากฏว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ ปรากฏว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้นจึงเปรียบเทียบ 2 กรณี คือ (1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ และ (2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว ผลปรากฏว่า วิธีชิปเทสท์ปรับใหม่ วิธีชิปเทสท์ และวิธีแมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี ส่วนวิธีการถดถอยโลจิสติกมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ปัจจัยทางด้านขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีดังกล่าว ดังนั้นจึงเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 และ 500 : 500 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่างคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

กรณีที่ 2 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างระดับเดียวกัน 6 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบทางเดียว พบว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีภายใต้ขนาดกลุ่มตัวอย่างทั้ง 6 ระดับมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Tukey ผลปรากฏว่า

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 500, 500 : 250, 1000 : 250 และ 1000 : 1000 ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าไม่แตกต่างกัน และทั้ง 2 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 500 อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับวิธีซิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

2.5 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธี กับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เงื่อนไขของปัจจัยที่ศึกษาปรากฏผลดังนี้

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% เมื่อเปรียบเทียบกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ภายในเกณฑ์เกือบทุกเงื่อนไขของลักษณะของข้อสอบ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของอีก 3 วิธีมีค่าอยู่ภายในเกณฑ์ทุกเงื่อนไขของลักษณะของข้อสอบ

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% เมื่อเปรียบเทียบกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ทุกเงื่อนไขของความยาวของแบบสอบ

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% เมื่อเปรียบเทียบกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

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% เมื่อเปรียบเทียบกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าอยู่ภายในเกณฑ์เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของอีก 3 วิธีมีค่าอยู่ภายในเกณฑ์ทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

จากผลดังกล่าว เมื่อพิจารณาในทุกเงื่อนไขของปัจจัยที่ศึกษา ปรากฏว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ที่ระดับ 10% โดยค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าประมาณ 8.88, 4.31, 4.90 และ 7.84 ตามลำดับ

อภิปรายผลการวิจัย

1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

1.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ พบว่า ลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อลักษณะของข้อสอบ a_L, b_M จะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด เมื่อลักษณะของข้อสอบ a_M, b_M จะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าสูงสุด เมื่อลักษณะของข้อสอบ a_H, b_M จะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงสุด ทั้งนี้มีข้อสังเกตว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นในทุกเงื่อนไขของลักษณะของข้อสอบที่มีค่าความยากปานกลาง (a_L, b_M; a_M, b_M และ a_H, b_M) แต่อำนาจการทดสอบของวิธีซิปเทสท์มีค่าเพิ่มขึ้นเพียงบางเงื่อนไขของลักษณะของข้อสอบที่มีค่าความยากปานกลาง (a_M, b_M) สาเหตุดังกล่าวอาจเนื่องมาจากข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมี 2 ลักษณะ คือ ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF) และข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (nondirecticnal DIF) (Li and Stout, 1993 cited in Narayanan and Swaminathan, 1996) ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวเกิดขึ้นเมื่อมีปฏิสัมพันธ์เป็นลำดับ(ordinal interaction) ระหว่างการเป็นสมาชิกของกลุ่มผู้สอบและระดับความสามารถ ซึ่งในทฤษฎีการตอบสนองข้อสอบ (item response theory; IRT) สามารถพิจารณาได้จากโค้งลักษณะข้อสอบ (item characteristic curves; ICCs) ระหว่างกลุ่มผู้สอบ 2 กลุ่มตัดกันตรงปลายสุดของช่วงความสามารถต่ำหรือสูง ข้อสอบลักษณะดังกล่าวจะมีค่าความยากต่ำหรือสูงซึ่งไม่มีปัญหาเมื่อตรวจสอบด้วยวิธีซิปเทสท์ ส่วนข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทางเกิดขึ้นเมื่อมีปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) ระหว่างการเป็นสมาชิกของกลุ่มผู้สอบและระดับความสามารถ ซึ่งทำให้โค้งลักษณะข้อสอบระหว่างกลุ่มผู้สอบ 2 กลุ่มตัดกันตรงจุดกึ่งกลางของช่วงความสามารถ ข้อสอบที่มีลักษณะนี้จะมีค่าความยากปานกลางซึ่งไม่สามารถตรวจสอบด้วยวิธีซิปเทสท์ ทั้งนี้เพราะว่าวิธีซิปเทสท์ใช้สถิติชนิดคิดเครื่องหมายคำนวณดัชนีข้อสอบที่ทำหน้าที่ต่างกัน เมื่อขนาดของข้อสอบที่ทำหน้าที่ต่างกันเปลี่ยนทิศทางตรงจุดกึ่งกลางของช่วงความสามารถจะทำให้ความแตกต่างที่มีเครื่องหมาย

ลบของช่วงคะแนนส่วนหนึ่งหักล้างกับความแตกต่างที่มีเครื่องหมายบวกของช่วงคะแนนอีกส่วนหนึ่ง ข้อสอบที่มีลักษณะดังกล่าวจึงไม่สามารถตรวจสอบด้วยวิธีซิปเทสท์ ดังนั้นในการศึกษาครั้งนี้ผู้วิจัยจึงนำวิธีซิปเทสท์ของ Shealy และ Stout (1993) มาปรับปรุงขั้นตอนการวิเคราะห์โดยแบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่มตามระดับคะแนนรวม คือ กลุ่มที่มีคะแนนสูงกับกลุ่มที่มีคะแนนต่ำ แล้ววิเคราะห์แยกกันในแต่ละกลุ่ม ผลการแบ่งกลุ่มผู้สอบจะทำให้โค้งลักษณะข้อสอบที่ตัดกันตรงจุดกึ่งกลางของช่วงความสามารถเปลี่ยนเป็นตัดกันตรงปลายสุดของช่วงความสามารถต่ำหรือสูง ดังนั้นจึงไม่มีผลต่อสถิติซนิดคิดเครื่องหมายของวิธีซิปเทสท์ปรับใหม่ นั่นคือ วิธีซิปเทสท์ปรับใหม่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปที่ไม่มีทิศทางได้ ส่วนวิธีแมนเทล-แฮนส์เซลก็สามารถอธิบายได้ในทำนองเดียวกัน ทั้งนี้เพราะว่าผู้วิจัยใช้วิธีแมนเทล-แฮนส์เซลแบบปรับปรุงขั้นตอนการวิเคราะห์ของ Mazor และคณะ (1994) ซึ่งใช้วิธีแบ่งกลุ่มผู้สอบ ดังนั้นจึงสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในลักษณะของข้อสอบที่มีค่าความยากปานกลางได้ทุกเงื่อนไข สำหรับวิธีการถดถอยโลจิสติกจะไม่มีปัญหาในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป เนื่องจากเป็นวิธีที่ใช้โมเดลการถดถอยโลจิสติกวิเคราะห์ดัชนีการทำหน้าที่ต่างกันของข้อสอบ โมเดลดังกล่าวสามารถคำนวณปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มและระดับความสามารถ ดังนั้นจึงทำให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปที่มีทิศทางเดียวและไม่มีทิศทางได้ (Swaminathan and Rogers, 1990) ผลการศึกษาในครั้งนี้ไม่สอดคล้องกับผลการศึกษาของ Rogers และ Swaminathan (1993) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าเพิ่มขึ้นเมื่อลักษณะของข้อสอบมีค่าความยากสูงขึ้นหรือต่ำลง ทั้งยังไม่สอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นเมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกสูงขึ้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอย-โลจิสติก ภายใต้ลักษณะของข้อสอบ 9 ลักษณะ พบว่า ภายใต้ลักษณะของข้อสอบ a_L, b_L และ a_L, b_H อำนาจการทดสอบระหว่างวิธีการตรวจสอบ 4 วิธีดังกล่าวมีค่าไม่แตกต่างกัน ผลการศึกษาดังกล่าวไม่สอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อลักษณะของข้อสอบที่มีค่าอำนาจจำแนกต่ำและมีค่าความยากต่ำหรือสูง เมื่อใช้วิธีการตรวจสอบ 4 วิธีดังกล่าวสามารถตรวจสอบได้ถูกต้องแม่นยำใกล้เคียงกัน ส่วนการตรวจสอบภายใต้ลักษณะของข้อสอบ 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 พบว่า ภายใต้เกือบทุกเงื่อนไขของการตรวจสอบ อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล โดยเฉพาะภายใต้ลักษณะของข้อสอบที่มีค่าอำนาจจำแนกต่ำกับความยากปานกลางจะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด ในขณะที่ภายใต้ลักษณะของข้อสอบที่มีค่าอำนาจจำแนกสูงกับความยากต่ำจะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์มีค่าต่ำสุด ผลการศึกษาดังกล่าวเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า วิธีซิปเทสท์ปรับใหม่ที่ผู้วิจัยนำมาปรับปรุงขั้นตอนการวิเคราะห์สามารถนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปได้อย่างมีประสิทธิภาพ โดยมีความถูกต้องแม่นยำในการตรวจสอบสูงกว่าวิธีซิปเทสท์แบบเดิม ทั้งยังมีความถูกต้องแม่นยำในการตรวจสอบเทียบเท่ากับวิธีการถดถอยโลจิสติก ทั้งนี้อาจเนื่องมาจากการปรับปรุงขั้นตอนการวิเคราะห์ของวิธีซิปเทสท์ซึ่งได้กล่าวมาแล้วข้างต้น จึงมีผลทำให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในลักษณะต่าง ๆ ได้ สำหรับวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า วิธีแมนเทล-แฮนส์เซลแบบปรับปรุงขั้นตอนการวิเคราะห์ถึงแม้ว่าจะมีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปสูงกว่าวิธีแมนเทล-แฮนส์เซลแบบเดิม (Mazor and others, 1994) แต่เมื่อนำมาเปรียบเทียบกับวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติก ปรากฏว่า มีอำนาจการทดสอบต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติก และเมื่อนำมาเปรียบเทียบกับวิธีซิปเทสท์ ปรากฏว่า มีอำนาจการทดสอบเท่าเทียมกัน ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของ Rogers และ Swaminathan (1993) Narayanan และ Swaminathan (1996) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป วิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไขของลักษณะของข้อสอบ แต่ไม่สอดคล้องกับผลการศึกษาของรัชรินทร์ มุคคา (2540) ที่พบว่า วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปเท่าเทียมกัน

1.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ ปรากฏว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก แสดงว่า

ในการตรวจสอบด้วย 4 วิธีดังกล่าว ถ้าใช้ความยาวของแบบสอบ 30 ข้อ จะมีความถูกต้องแม่นยำใกล้เคียงกับการตรวจสอบโดยใช้ความยาวของแบบสอบ 60 ข้อ นั่นคือ ถึงแม้ว่าความยาวของแบบสอบจะเพิ่มขึ้น ซึ่งจะส่งผลให้ความเที่ยงของคะแนนเพิ่มขึ้นด้วย แต่จะไม่มีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 4 วิธี ผลการศึกษาดังกล่าวมีทั้งสอดคล้องและไม่สอดคล้องกับผลการศึกษาของ Swaminathan และ Rogers (1990) Rogers และ Swaminathan (1993) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซล แต่จะมีผลต่อวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อความยาวของแบบสอบเพิ่มขึ้นจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นด้วย ดังนั้นในกรณีของวิธีแมนเทล-แฮนส์เซลจะสอดคล้อง ส่วนในกรณีของวิธีการถดถอยโลจิสติกจะไม่สอดคล้อง นอกจากนี้ยังไม่สอดคล้องกับผลการศึกษาของ Ackerman และ Evans (1992 cited in Narayanan and Swaminathan, 1994) ที่ให้ข้อเสนอแนะไว้ว่า เมื่อแบบสอบมีความเที่ยงเพิ่มขึ้นในระดับปานกลางจะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าเพิ่มขึ้น แต่เมื่อแบบสอบมีความเที่ยงต่ำหรือสูงจะมีผลทำให้อำนาจการทดสอบมีค่าลดลง

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ความยาวของแบบสอบ 30 ข้อและ 60 ข้อ ปรากฏว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขของการตรวจสอบ อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ผลการศึกษาดังกล่าวเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 1 สาเหตุดังกล่าวมีลักษณะทำนองเดียวกับที่กล่าวไว้ในข้อ 1.1 ผลการศึกษาในครั้งนี้สอดคล้องกับผลการศึกษาของจิติมา วรรณศรี (2539) ที่พบว่า อำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไขของความยาวของแบบสอบ นอกจากนี้ยังสอดคล้องกับผลการศึกษาของ Swaminathan และ Rogers (1990) Rogers และ Swaminathan (1993) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ทุกเงื่อนไขของความยาวของแบบสอบ

1.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน 3 ระดับ ปรากฏว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ แต่จะมีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลงจาก 20% เป็น 5% จะมีผลทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น สาเหตุดังกล่าวอาจเนื่องมาจากเมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง จะมีผลทำให้ค่าความเที่ยงของคะแนนรวมมีค่าเพิ่มขึ้น และเมื่อความเที่ยงของคะแนนรวมมีค่าเพิ่มขึ้นจะมีผลทำให้ค่าประมาณความสามารถที่ใช้ในการทำนายด้วยโมเดลการถดถอยโลจิสติกและใช้เป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบของวิธีแมนเทล-แฮนส์เซลมีความถูกต้องแม่นยำเพิ่มขึ้น (Swaminathan and Rogers, 1990) ดังนั้นจึงมีผลทำให้อำนาจการทดสอบของ 2 วิธีดังกล่าวมีค่าเพิ่มขึ้น ผลการศึกษาดังกล่าวเฉพาะวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกไม่สอดคล้องกับผลการศึกษาของ Rogers และ Swaminathan (1993) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก นอกจากนี้ยังมีทั้งสอดคล้องและไม่สอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีโคร-ซิปและวิธีแมนเทล-แฮนส์เซล แต่จะมีผลต่ออำนาจการทดสอบวิธีการถดถอยโลจิสติก โดยมีอำนาจการทดสอบเพิ่มขึ้นเมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง นั่นคือ ในกรณีของวิธีการถดถอยโลจิสติกจะได้ผลสอดคล้องกัน ส่วนในกรณีของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์จะได้ผลสอดคล้องกับวิธีโคร-ซิป แต่ในกรณีของวิธีแมนเทล-แฮนส์เซลได้ผลไม่สอดคล้องกัน จากผลการศึกษาดังกล่าวมีข้อสังเกตว่า ในกรณีของวิธีแมนเทล-แฮนส์เซลจะไม่สอดคล้องกับผลการศึกษาของคนอื่น ๆ เสมอ ทั้งนี้อาจเนื่องมาจากผู้วิจัยใช้วิธีแมนเทล-แฮนส์เซลของ Mazor และคณะ (1994) แทนวิธีแมนเทล-แฮนส์เซลของ Holland และ Thayer (1988) จึงทำให้ผลการตรวจสอบไม่สอดคล้องกัน ผลการวิจัยในครั้งนี้มีข้อค้นพบที่น่าสนใจเกี่ยวกับสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน กล่าวคือ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ ดังนั้นในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วย 2 วิธีดังกล่าว สามารถปรับปรุงการวิเคราะห์เป็นแบบ 2 ขั้นตอน เพื่อทำให้ผลการวิเคราะห์มีประสิทธิภาพมากยิ่งขึ้น (Rogers และ Swaminathan, 1993; Narayanan and Swaminathan, 1994) โดยในขั้นตอนแรก

จะคำนวณคะแนนรวมของผู้สอบจากข้อสอบทุกข้อเพื่อใช้เป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบแล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ และในขั้นตอนที่สองจะนำข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันออกจากการคำนวณคะแนนรวมของผู้สอบ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง ผลที่ได้จะทำให้เกณฑ์การจับคู่กลุ่มผู้สอบมีความบริสุทธิ์ (purification) ซึ่งจะมีผลทำให้การตรวจสอบมีความถูกต้องแม่นยำสูง

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% ปรากฏว่า ภายใต้เกือบทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน อำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ผลการศึกษาดังกล่าวเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 1 สาเหตุดังกล่าวมีลักษณะทำนองเดียวกับที่กล่าวไว้ในข้อ 1.1 นอกจากนี้ยังพบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% จะส่งผลให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ ส่วนภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะส่งผลให้อำนาจการทดสอบของวิธีซิปเทสท์มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการตรวจสอบ ทั้งนี้อาจเนื่องมาจากภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5% จะมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด ซึ่งจะส่งผลให้ความเที่ยงของคะแนนรวมมีค่าสูงสุด ดังนั้นค่าประมาณความสามารถที่ใช้ในการทำนายด้วยโมเดลการถดถอยโลจิสติกจะมีความถูกต้องแม่นยำสูง นั่นคือ ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติกจะมีอำนาจการทดสอบสูง ในขณะเดียวกัน ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ซึ่งจะส่งผลให้ความเที่ยงของคะแนนรวมมีค่าต่ำสุด ดังนั้นคะแนนรวมที่ใช้เป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบของวิธีซิปเทสท์จะมีความถูกต้องแม่นยำต่ำ นั่นคือ ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีซิปเทสท์จะมีอำนาจการทดสอบต่ำ

1.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ขนาดกลุ่มตัวอย่างที่ต่างกัน 6 ระดับ พบว่า ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจาก $N_R : N_F$ = 250 : 250 และ 500 : 250 เป็น $N_R : N_F$ = 1000 : 1000 จะมีผลทำให้อำนาจการทดสอบของทั้ง 4 วิธีมีค่าเพิ่มขึ้น สาเหตุดังกล่าวอาจเนื่องมาจากค่าประมาณดัชนีการทำหน้าที่ต่างกันของข้อสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซล คำนวณโดยใช้สัดส่วนการตอบข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าใช้กลุ่มตัวอย่างขนาดเล็กจะมีผลต่อความคงที่ (stability) ของค่าประมาณดัชนีในแต่ละกลุ่มผู้สอบ ส่วนวิธีการถดถอยโลจิสติกถ้าใช้กลุ่มตัวอย่างขนาดเล็กจะมีผลทำให้การทดสอบสถิติไค-สแควร์ที่มีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic) ขาดความตรง (Swaminathan and Rogers, 1990) ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของจิติมา วรรณศรี (2539) Mazor และคณะ (1992) Rogers และ Swaminathan (1993) Narayanan และ Swaminathan (1994, 1996) ที่พบว่า เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล วิธีซิปเทสท์ วิธีโคร-ซิป และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น นอกจากนี้ยังพบว่า อำนาจการทดสอบของทั้ง 4 วิธีภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 มีค่าไม่แตกต่างกัน แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 1000 : 1000 จะทำให้ผลการตรวจสอบมีความถูกต้องแม่นยำสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 แต่ถ้าใช้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250 จะทำให้ผลการตรวจสอบมีความถูกต้องแม่นยำเท่ากับขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 นั่นคือ การตรวจสอบด้วยวิธีทั้ง 4 เมื่อต้องการใช้ขนาดกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบต่ำกว่า 1000 คน ถ้าใช้เพียงกลุ่มละ 250 คนก็มีประสิทธิภาพเพียงพอในการตรวจสอบแล้ว ผลการศึกษาดังกล่าวสอดคล้องกับข้อเสนอแนะของ Hill (1990 cited in Mazor and others, 1992) ที่กล่าวว่า การตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซลสามารถใช้ขนาดกลุ่มตัวอย่างในแต่ละกลุ่มหรือทั้ง 2 กลุ่ม ระหว่าง 100 ถึง 300 คน ผลการศึกษาครั้งนี้มีข้อสังเกตที่น่าสนใจเกี่ยวกับอัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีผลต่ออำนาจการทดสอบ กล่าวคือ ในทุกเงื่อนไขของการตรวจสอบ เมื่อเพิ่มขนาดผู้สอบกลุ่มเปรียบเทียบจะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธี

การถดถอยโลจิสติกมีค่าเพิ่มขึ้นมากกว่าเพิ่มขนาดผู้สอบกลุ่มอ้างอิง ดังเช่น เมื่อ N_F = 250 และ N_R เพิ่มจาก 250 เป็น 1000 แล้วอำนาจการทดสอบของทั้ง 4 วิธีดังกล่าวมีค่าเพิ่มขึ้นประมาณร้อยละ 4, 4, 3 และ 0 ตามลำดับ แต่เมื่อ N_R = 1000 และ N_F เพิ่มจาก 250 เป็น 1000 อำนาจการทดสอบของทั้ง 4 วิธีดังกล่าวมีค่าเพิ่มขึ้นประมาณร้อยละ 18, 9, 13 และ 22 ตามลำดับ จะเห็นว่าอำนาจการทดสอบของทั้ง 4 วิธีมีค่าเพิ่มขึ้น ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1994) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์ เมื่อเพิ่มขนาดผู้สอบกลุ่มเปรียบเทียบจะมีผลทำให้อำนาจการทดสอบของทั้ง 2 วิธีดังกล่าวมีค่าเพิ่มขึ้นมากกว่าเพิ่มขนาดผู้สอบกลุ่มอ้างอิง ทั้งยังสอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่ศึกษาในกรณีแบบเอกรูปของวิธีโคร-ซิป วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยได้ผลในลักษณะเดียวกัน

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอย-โลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง N_R : N_F = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000 ปรากฏว่า ภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ผลการศึกษาดังกล่าวเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 1 สาเหตุดังกล่าวมีลักษณะเดียวกับที่กล่าวไว้ในข้อ 1.1 ผลการศึกษาในครั้งนี้สอดคล้องกับผลการศึกษาของ Rogers และ Swaminathan (1993) Narayanan และ Swaminathan (1996) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลภายใต้ทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

2. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบต่างกัน 9 ลักษณะ พบว่า ลักษณะของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อลักษณะของข้อสอบ a_M, b_H จะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงสุด เมื่อลักษณะของข้อสอบ a_L, b_H จะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าสูงสุด และเมื่อลักษณะของข้อสอบ a_L, b_L จะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงสุด ทั้งนี้มีข้อสังเกตว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นในทุกเงื่อนไขของลักษณะของข้อสอบที่มีค่าอำนาจจำแนกลดลง (a_L, b_L; a_L, b_M และ a_L, b_H) ผลการศึกษาดังกล่าวไม่สอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกเพิ่มขึ้นจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นภายใต้เกือบทุกเงื่อนไขของลักษณะของข้อสอบ

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้ลักษณะของข้อสอบ 9 ลักษณะ ปรากฏว่า ภายใต้เกือบทุกเงื่อนไขของลักษณะของข้อสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ผลการศึกษาดังกล่าวเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 สาเหตุดังกล่าวอาจเนื่องมาจากในการศึกษาครั้งนี้ผู้วิจัยจำลองข้อมูลโดยกำหนดให้การแจกแจงค่าความสามารถของกลุ่มผู้สอบเป็นแบบปกติ (มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1) เหมือนกันในทุกเงื่อนไขของการจำลองข้อมูล ต่อจากนั้นจึงจัดกระทำขนาดกลุ่มตัวอย่างระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบโดยการสุ่ม ผลจากการสุ่มจะทำให้การแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเท่ากัน

ซึ่งจะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 แต่ถ้าการแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าแตกต่างกันจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้น (Rogers and Swaminathan, 1993; Narayanan and Swaminathan, 1996; Roussos and Stout, 1996) ในการวิเคราะห์ด้วยวิธีซิปเทสท์ผู้วิจัยใช้ข้อมูลที่ได้จากการสุ่มดังที่กล่าวมาข้างต้น ดังนั้นจึงไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ แต่การวิเคราะห์ด้วยวิธีซิปเทสท์ปรับใหม่ ผู้วิจัยปรับปรุงขั้นตอนการวิเคราะห์โดยแบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่มตามระดับคะแนนรวม คือ กลุ่มที่มีคะแนนสูงกับกลุ่มที่มีคะแนนต่ำ แล้ววิเคราะห์แยกกันในแต่ละกลุ่ม ผลจากการแบ่งกลุ่มดังกล่าวอาจมีผลทำให้การแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าแตกต่างกัน ซึ่งจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าเพิ่มขึ้น ดังนั้นวิธีซิปเทสท์ปรับใหม่จึงมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ปรากฏว่า มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ ทั้งนี้เนื่องจากวิธีซิปเทสท์มีรูปแบบนันพาราเมทริก (nonparametric) เมื่อทดสอบกับสถิติ Z ซึ่งมีการแจกแจงแบบปกติไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ แต่วิธีการถดถอยโลจิสติกมีรูปแบบพาราเมทริก (parametric) เมื่อทดสอบกับสถิติไค-สแควร์ซึ่งมีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic) จะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ถ้าใช้กลุ่มตัวอย่างขนาดเล็กจะทำให้การทดสอบขาดความตรง (Swaminathan and Rogers, 1990) ซึ่งจะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้น ในการศึกษาครั้งนี้ผู้วิจัยใช้ขนาดกลุ่มตัวอย่างหลายระดับ ตั้งแต่ขนาดเล็ก (N_R : N_F = 250 : 250) จนถึงขนาดใหญ่ (N_R : N_F = 1000 : 1000) ดังนั้นในกลุ่มตัวอย่างขนาดเล็กจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 2 สาเหตุดังกล่าวอาจเนื่องมาจากในการศึกษาครั้งนี้ ผู้วิจัยใช้การวิเคราะห์แบบ 2 ขั้นตอน กล่าวคือ ในขั้นตอนแรกจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์สำหรับการจับคู่ของกลุ่มผู้สอบ และในขั้นตอนที่สองจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง แต่จะนำข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในขั้นตอนแรกออกก่อนที่จะคำนวณคะแนนรวมของผู้สอบ การวิเคราะห์ดังกล่าวจะทำให้เกณฑ์การจับคู่ของกลุ่มผู้สอบมีความบริสุทธิ์ ซึ่งจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง ดังนั้นวิธีแมนเทล-แฮนส์เซลจึงมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติก แต่เมื่อเปรียบเทียบกับวิธีซิปเทสท์ ปรากฏว่า มีค่าไม่แตกต่างกัน นอกจากนี้ยังพบว่า ภายใต้ลักษณะ

ของข้อสอบ a ต่ำกับ b ปานกลางจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์-ปรับใหม่มีค่าสูงสุด ทั้งนี้อาจเนื่องมาจากภายใต้ลักษณะของข้อสอบดังกล่าว อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่มีค่าสูงสุด ซึ่งโดยทั่ว ๆ ไปแล้วค่าความคลาดเคลื่อนประเภทที่ 1 ก็คือค่า α นั่นเอง ซึ่งจะแปรผันตามค่าอำนาจการทดสอบ เมื่อวิธีซิปเทสท์ปรับใหม่มีอำนาจการทดสอบสูงสุดภายใต้ลักษณะของข้อสอบ a ต่ำกับ b ปานกลางจะส่งผลให้มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดด้วย ผลการศึกษาในครั้งนี้สอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้ทุกเงื่อนไขของลักษณะของข้อสอบ

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ความยาวของแบบสอบต่างกัน 2 ระดับ พบว่า ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก แต่จะมีผลต่ออัตราความ-คลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์ กล่าวคือ เมื่อความยาวของแบบสอบลดลงจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์มีค่าเพิ่มขึ้น สาเหตุดังกล่าวอาจเนื่องมาจากแบบสอบที่มีความยาวลดลงจะมีผลทำให้ความเที่ยงของค่าประมาณความสามารถมีค่าลดลง เมื่อความเที่ยงของค่าประมาณความสามารถลดลงจะส่งผลให้ความถูกต้องแม่นยำในการจับคู่ของกลุ่มผู้สอบมีค่าลดลงด้วย (Narayanan and Swaminathan, 1994) ดังนั้นอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีซิปเทสท์จึงมีค่าเพิ่มขึ้น ผลการศึกษาดังกล่าวมีทั้งสอดคล้องและไม่สอดคล้องกับผลการศึกษาของ Swaminathan และ Rogers (1990) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล แต่จะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก โดยจะมีค่าเพิ่มขึ้นเมื่อเพิ่มความยาวของแบบสอบ ดังนั้นวิธีแมนเทล-แฮนส์เซลจะสอดคล้อง ส่วนวิธีการถดถอยโลจิสติกจะไม่สอดคล้องกับผลการศึกษาดังกล่าว นอกจากนี้วิธีแมนเทล-แฮนส์เซลยังไม่สอดคล้องกับผลการศึกษาของ Uttaro และ Millsap (1994) ที่พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้แบบสอบที่มีขนาดยาวจะมีค่าสูงกว่าอัตราความ-คลาดเคลื่อนประเภทที่ 1 ภายใต้แบบสอบที่มีขนาดสั้น

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้ความยาวของแบบสอบ 30 ข้อ และ 60 ข้อ ปรากฏว่า ได้ผลสอดคล้องกันกล่าวคือ ภายใต้ทุกเงื่อนไขของการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ผลการศึกษาเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่หนดไว้ในข้อ 2 สาเหตุดังกล่าวมีลักษณะทำนองเดียวกับที่กล่าวไว้ในข้อ 2.1 ผลการศึกษาในครั้งนี้สอดคล้องกับผลการศึกษาของ Swaminathan และ Rogers (1990) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไขของความยาวของแบบสอบ

2.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนต่างกัน 3 ระดับ ปรากฏว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก แต่จะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ กล่าวคือ เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนเพิ่มขึ้นจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าเพิ่มขึ้น สาเหตุดังกล่าวอาจเนื่องมาจากสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนเพิ่มมากขึ้นจะมีผลทำให้ความเที่ยงของค่าประมาณความสามารถมีค่าลดลง เมื่อค่าประมาณความสามารถมีความเที่ยงลดลงจะส่งผลให้ความแม่นยำในการจับคู่ของกลุ่มผู้สอบมีค่าลดลงด้วย (Narayanan and Swaminathan, 1994) ดังนั้นอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่จึงมีค่าเพิ่มขึ้น ผลดังกล่าวสอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก แต่จะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโคร-ซิป โดยจะมีค่าเพิ่มขึ้นเมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนเพิ่มขึ้น

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% พบว่า ได้ผลสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขของการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน และทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล นอกจากนี้ยังพบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าต่ำสุด ทั้งนี้อาจเนื่องมาจากภายใต้ลักษณะของข้อสอบดังกล่าว อำนาจการทดสอบของวิธีซิปเทสท์มีค่าต่ำสุด ซึ่งโดยทั่ว ๆ ไปแล้ว ค่าความคลาดเคลื่อนประเภทที่ 1 (α) จะแปรผันตามค่าอำนาจการทดสอบ เมื่อวิธีซิปเทสท์มีอำนาจการทดสอบต่ำสุดภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% จะส่งผลให้มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุดด้วย ผลการศึกษาในครั้งนี้เฉพาะวิธีซิปเทสท์-ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่หนดไว้ในข้อ 2 สาเหตุดังกล่าวมีลักษณะเดียวกับที่กล่าวไว้ในข้อ 2.1 ผลการศึกษาในครั้งนี้สอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1996) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลภายใต้ทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ

2.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 6 ระดับ ปรากฏว่า ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซล แต่จะมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อกลุ่มตัวอย่างมีขนาดลดลงจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น ทั้งนี้อาจเนื่องมาจากการทดสอบทางสถิติของวิธีการถดถอยโลจิสติกจะใช้สถิติไค-สแควร์ที่มีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic) เมื่อใช้กลุ่มตัวอย่างขนาดเล็กจะมีผลทำให้การทดสอบดังกล่าวขาดความตรง (Swaminathan and Rogers, 1990) ดังนั้นอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกจึงมีค่าเพิ่มขึ้น ผลการศึกษาดังกล่าวไม่สอดคล้องกับผลการศึกษาของ

Narayanan และ Swaminathan (1996) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโคร-ซิป วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250, 1000 : 500 และ 1000 : 1000 ปรากฏว่า ได้ผลสอดคล้องกัน กล่าวคือ ภายใต้เกือบทุกเงื่อนไขของการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกัน ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล ผลการศึกษาดังกล่าวเฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 2 สาเหตุดังกล่าวมีลักษณะทำนองเดียวกับที่กล่าวไว้ในข้อ 2.1 ผลการศึกษาในครั้งนี้สอดคล้องกับผลการศึกษาของ Swaminathan และ Rogers (1990) Narayanan และ Swaminathan (1996) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล ภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

จากผลการศึกษาทั้งหมดจะเห็นว่า เมื่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของ 2 วิธีดังกล่าวจะมีค่าสูงด้วย ในขณะเดียวกัน เมื่ออำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของ 2 วิธีดังกล่าวจะมีค่าต่ำด้วย แสดงว่า ค่าทั้งสองแปรผันตามกัน ซึ่งเป็นไปตามทฤษฎีของค่าอำนาจการทดสอบและค่าความคลาดเคลื่อนประเภทที่ 1 (α) อย่างไรก็ตาม วิธีการตรวจสอบที่มีประสิทธิภาพสูงควรมีอำนาจการทดสอบสูงและอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ ดังนั้นจึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 4 วิธีไปเปรียบเทียบกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ผลปรากฏว่า ภายใต้เงื่อนไขของปัจจัยที่ศึกษาทั้งหมด 324 เงื่อนไข อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าประมาณ

8.88%, 4.31%, 4.90% และ 7.84% ตามลำดับ ผลดังกล่าวจะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ที่กำหนด โดยมีค่าเฉลี่ยต่ำกว่า 10% ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานในข้อ 2 และสอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1994, 1996) ที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป ผลการตรวจสอบแบบเอกรูป พบว่า ภายใต้ทุกเงื่อนไขของการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าเฉลี่ย 4.75% และ 6.65% ตามลำดับ ส่วนผลการตรวจสอบแบบอเนกรูป พบว่า ภายใต้ทุกเงื่อนไขของการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล วิธีโคร-ซิป และวิธีการถดถอยโลจิสติกมีค่าเฉลี่ย 4.81%, 8.45% และ 8.15% ตามลำดับ จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของทุกวิธีการตรวจสอบมีค่าเฉลี่ยต่ำกว่า 10%

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 จากผลการวิจัย พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ภายใต้ปัจจัยที่ศึกษา 4 ตัว คือ ลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันและขนาดกลุ่มตัวอย่าง อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ทั้งวิธีซิปเทสท์-ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไข เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ปัจจัยที่ศึกษา 324 เงื่อนไขกับเกณฑ์ที่ระดับ 10% พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ดังกล่าว นั่นคือ วิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปได้เท่าเทียมกัน และทั้ง 2 วิธีมีความถูกต้องแม่นยำสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล โดยทั่ว ๆ ไปแล้วนักวิธีวิทยาการวิจัยได้ให้การยอมรับการตรวจสอบด้วยวิธีการถดถอยโลจิสติก ทั้งนี้เนื่องจากวิธีดังกล่าวใช้โมเดลการถดถอยโลจิสติกซึ่งสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป แต่วิธีดังกล่าวยังมีข้อจำกัดหลายประการ ดังเช่น การประมาณค่าพารามิเตอร์ในโมเดลการถดถอยโลจิสติกมีความ-

ซับซ้อนเพราะต้องคำนวณทวนซ้ำหลายรอบ (iterative) ซึ่งทำให้การคำนวณต้องใช้เวลามาก ทั้งยังเสียค่าใช้จ่ายค่อนข้างสูง นอกจากนี้การใช้สถิติไค-สแควร์ที่มีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic) ทดสอบนัยสำคัญจะต้องใช้ขนาดกลุ่มตัวอย่างที่เพียงพอ เพื่อมิให้การทดสอบขาดความตรง ส่วนวิธีซิปเทสท์ปรับใหม่มีข้อได้เปรียบตรงที่สามารถคำนวณได้ง่ายไม่ซับซ้อน โดยประมาณค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบจากสัดส่วนของการตอบข้อสอบระหว่างกลุ่มผู้สอบ 2 กลุ่ม ไม่ต้องคำนวณทวนซ้ำหลายรอบ ซึ่งทำให้คำนวณได้เร็ว ทั้งยังไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทำให้เสียค่าใช้จ่ายไม่มาก นอกจากนี้ยังตัดสินข้อสอบที่ทำหน้าที่ต่างกันโดยใช้สถิติ Z ทดสอบนัยสำคัญทางสถิติทำให้มีความถูกต้องสูง ดังนั้นวิธีซิปเทสท์ปรับใหม่จึงเป็นอีกทางเลือกหนึ่งที่สามารถนำมาใช้แทนวิธีการถดถอยโลจิสติกได้

1.2 จากผลการวิจัย พบว่า ลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อลักษณะของข้อสอบมีค่าความยากปานกลางจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น และเมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกปานกลางถึงสูงจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีดังกล่าวมีค่าลดลง แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบที่มีค่าความยากปานกลางและอำนาจจำแนกปานกลางถึงสูง วิธีการถดถอยโลจิสติกสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องแม่นยำ หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้น้อยมาก จากผลการศึกษาดังกล่าวสามารถนำวิธีการถดถอยโลจิสติกไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในแบบสอบวัดผลสัมฤทธิ์ทางการเรียน ทั้งนี้เนื่องจากในแบบสอบดังกล่าวมักจะมีข้อสอบที่มีค่าความยากปานกลางและอำนาจจำแนกสูง ดังนั้นถ้าตรวจสอบด้วยวิธีการถดถอยโลจิสติกจะส่งผลให้การตรวจสอบมีประสิทธิภาพสูง

1.3 จากผลการวิจัย พบว่า ความยาวของแบบสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับ-ใหม่ กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบสอบ 60 ข้อ มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบสอบ 30 ข้อ แสดงว่า เมื่อเพิ่มความยาวของแบบสอบจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง นั่นคือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในแบบสอบ 60 ข้อ ควรใช้วิธีซิปเทสท์ปรับใหม่ ทั้งนี้จะทำให้ระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้น้อยที่สุด

1.4 จากผลการวิจัย พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ กล่าวคือ อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีดังกล่าวภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% มีค่าใกล้เคียงกัน แสดงว่า เมื่อตรวจสอบด้วยวิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องใกล้เคียงกันหรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้ใกล้เคียงกัน จากข้อค้นพบดังกล่าวสามารถนำวิธีซิปเทสท์มาปรับปรุงขั้นตอนการวิเคราะห์โดยใช้วิธีการวิเคราะห์แบบ 2 ขั้นตอน กล่าวคือ ในขั้นตอนแรกจะคำนวณคะแนนรวมของผู้สอบจากข้อสอบทุกข้อเพื่อใช้เป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ในขั้นตอนที่สองจะนำข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันออกจากการคำนวณคะแนนรวมของผู้สอบ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง ผลการวิเคราะห์จะทำให้เกณฑ์ในการจับคู่ของกลุ่มผู้สอบมีความบริสุทธิ์ (purification) ซึ่งจะส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์มีประสิทธิภาพมากยิ่งขึ้น

1.5 จากผลการวิจัย พบว่า ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซล กล่าวคือ เมื่อใช้วิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1000 : 250 และ 1000 : 500 จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องใกล้เคียงกัน หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้ใกล้เคียงกัน ดังนั้น ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล เมื่อต้องการใช้ขนาดกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบต่ำกว่า 1000 คนสามารถใช้เพียงกลุ่มละ 250 คน ก็สามารถตรวจสอบได้อย่างมีประสิทธิภาพ

2. ข้อเสนอแนะในการวิจัยต่อไป

2.1 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป โดยศึกษาเฉพาะกรณีของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีทิศทางเดียว (unidirectional) และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่ไม่มีทิศทาง (non-unidirectional) ตามกรอบของทฤษฎี IRT ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีทิศทางเดียวสามารถพิจารณาได้จากโค้งลักษณะข้อสอบ (item characteristic curves) ระหว่างผู้สอบ 2 กลุ่มที่ไม่ขนานกันแต่ไม่ตัดกัน สำหรับข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่ไม่มีทิศทางสามารถพิจารณาได้จากโค้งลักษณะข้อสอบระหว่างผู้สอบ 2 กลุ่มที่ไม่ขนานกันและตัดกัน

2.2 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป โดยใช้ข้อมูลจริงเพื่อเปรียบเทียบกับการศึกษาในครั้งนี้ นอกจากนี้ควรศึกษาร่วมกับตัวแปรที่เกี่ยวกับลักษณะของผู้สอบ เช่น เพศ เชื้อชาติ ศาสนา ภูมิลำเนา อายุ และประสบการณ์ เป็นต้น

2.3 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ซึ่งมีลักษณะคล้ายกับการวิจัยในครั้งนี้ โดยศึกษาร่วมกับตัวแปรอื่น ๆ ที่คาดว่าน่าจะมีผลต่ออำนาจการทดสอบและอัตราความคาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ เช่น ตัวแปรความแตกต่างของการแจกแจงค่าความสามารถ (ability distribution differences) ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ตัวแปรขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกัน (DIF effect size) ตัวแปรเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ (matching criterion) และตัวแปรอัตราส่วน (ratio) ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ เป็นต้น นอกจากนี้ยังมีตัวแปรที่เกี่ยวกับวิธีการตรวจสอบในกลุ่ม IRT โดยเฉพาะ เช่น วิธีการประมาณค่าพารามิเตอร์ของข้อสอบและวิธีการปรับเทียบสเกลค่าพารามิเตอร์ของข้อสอบ เป็นต้น

2.4 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (polytomous) เช่น แบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่มีการให้คะแนนแบบบางส่วน มาตรวัดเจตคติ มาตรวัดบุคลิกภาพ การประเมินทักษะการปฏิบัติ และการประเมินตามสภาพที่แท้จริง เป็นต้น โดยใช้วิธีในกลุ่ม IRT ได้แก่ วิธีการทดสอบไค-สแควร์ของ Lord วิธีการทดสอบอัตราส่วนไลค์ลิฮู้ด วิธีการวัดพื้นที่ของ Raju วิธีการวัดพื้นที่ของ Kim และ Cohen วิธีซิปเทสท์ และวิธีซิปเทสท์ปรับใหม่ เป็นต้น หรือใช้วิธีในกลุ่ม non-IRT ได้แก่ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (logistic discriminant function analysis) วิธีการทำให้เป็นมาตรฐาน และวิธีแมนเทล-แฮนส์เซล เป็นต้น

2.5 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อมีผู้สอบหลายกลุ่ม (multiple groups) โดยใช้วิธีในกลุ่ม IRT และ non-IRT และควรศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (differential bundle functioning; DBF) ทั้งในแบบสอบเอกมิติ (unidimensional test) และแบบสอบพหุมิติ (multi-dimensional test) นอกจากนี้ควรศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (computerized adaptive testing; CAT)