

สรุปผลการวิจัยและอภิปรายผล

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบสถิติที่ใช้ทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม คือ การทดสอบเอฟ (F test) การทดสอบคลอทซ์ (Klotz test) การทดสอบมูด (Mood test) การทดสอบทูกี้-ซีเกล (Tukey-Siegel) และการทดสอบสแควร์เรงค์ (squared ranks test) โดยศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบทั้ง 5 วิธี ดังกล่าว เมื่อประชากรมีการแจกแจงแบบปกติ (Normal Distribution) แบบยูนิฟอร์ม (Uniform Distribution) และแบบโลจิสติก (Logistic Distribution) แต่ในกรณีที่ประชากรมีการแจกแจงแบบเบ้ (Skewed Distribution) จะศึกษาเปรียบเทียบเฉพาะการทดสอบเอฟ และการทดสอบสแควร์เรงค์ (\bar{X}) เท่านั้น ซึ่งในกรณีที่ประชากรมีการแจกแจงแบบเบ้ศึกษาอำนาจของการทดสอบและความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อความเบ้มีค่าเป็นบวกและมีค่าสูงที่สุด 1.0 และความโด่งมีค่าอยู่ในช่วง [2.8, 4.2] โดยสนใจศึกษาเมื่อขนาดของตัวอย่างทั้งสองกลุ่มมีค่าเท่ากัน คือ 10 15 และ 20 และขนาดของตัวอย่างทั้งสองกลุ่มมีค่าไม่เท่ากัน คือ ขนาดของตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 เป็นดังนี้ (5;10) (5;10 และ (5;20) ซึ่งผลสรุปจากการวิจัยจำแนกได้ดังนี้

ผลสรุปของความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1

จากการทดสอบหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบเอฟ การทดสอบคลอทซ์ การทดสอบมูด การทดสอบทูกี้-ซีเกล และการทดสอบสแควร์เรงค์ ซึ่งเมื่อนำมาเปรียบเทียบกับค่า α ที่กำหนด โดยใช้เกณฑ์พิจารณาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ของ Bradley และ Cochran จำแนกผลสรุปได้ดังนี้

1. ในกรณีที่ประชากรมีการแจกแจงแบบปกติ แบบยูนิฟอร์ม และแบบโลจิสติก สรุปผลได้ดังนี้

1.1 การทดสอบคล้อย การทดสอบมัด การทดสอบบูกี-ซีเกล และ การทดสอบสแควร์แรงค์ (μ) สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี โดยเฉพาะ เมื่อ $\alpha = .05$ และ $.10$ สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทั้งหมด เมื่อใช้ เกณฑ์ของ Bradley สำหรับการทดสอบเอฟควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีในกรณี ที่ประชากรมีการแจกแจงแบบปกติ แต่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้น้อยมากเมื่อ ประชากรมีการแจกแจงแบบยูนิฟอร์มและแบบโลจิสติก ไม่ว่าจะใช้เกณฑ์ของ Bradley หรือเกณฑ์ของ Cochran ซึ่งผลสรุปของความสามารถในการควบคุมความคลาดเคลื่อนประเภท ที่ 1 ของสถิติทดสอบทั้ง 5 วิธีดังกล่าว แสดงได้ดังตาราง 5.1 สำหรับลักษณะที่ควบคุมความ คลาดเคลื่อนประเภทที่ 1 ไม่ได้ของการทดสอบคล้อย การทดสอบมัด การทดสอบบูกี-ซีเกล และการทดสอบสแควร์แรงค์ (μ) นั้น ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองจะน้อยกว่าค่า α ที่กำหนด แต่ในกรณีของการทดสอบเอฟ ลักษณะที่ควบคุม ความคลาดเคลื่อนประเภทที่ 1 ไม่ได้ถือค่าความน่าจะเป็นของความคลาดเคลื่อนประเภท ที่ 1 จากการทดลองมากกว่าค่า α ที่กำหนด เมื่อประชากรมีการแจกแจงแบบโลจิสติก และ ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองน้อยกว่า α ที่กำหนด เมื่อประชากรมีการแจกแจงแบบยูนิฟอร์ม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตาราง 5.1 แสดงจำนวนครั้งที่การทดสอบเอฟ (F) การทดสอบคล็อยซ์ (KN) การทดสอบมุต (M) การทดสอบทูกี-ซีเกล (TS) และการทดสอบสแควร์แรงค์ (μ) (SRT(μ)) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ จากการทดลองทั้งหมด 6 กรณี สำหรับแต่ละรูปแบบของการแจกแจง ค่าแจกตามระดับนัยสำคัญที่กำหนด

ระดับ นัยสำคัญ	สถิติทดสอบ	1.เกณฑ์ของ Bradley			1.เกณฑ์ของ Cochran		
		NN	UU	LL	NN	UU	LL
.01	F	5	1	0	5	0	0
	KN	5	4	4	4	3	3
	M	6	6	6	6	5	5
	TS	6	6	6	5	6	6
	SRT(μ)	6	6	6	6	6	6
.05	F	6	0	1	5	0	0
	KN	6	6	6	6	5	5
	M	6	6	6	6	6	6
	TS	6	6	6	3	5	5
	SRT(μ)	6	6	6	6	3	3
.10	F	6	0	3	6	0	0
	KN	6	6	6	6	5	5
	M	6	6	6	6	6	6
	TS	6	6	6	6	6	6
	SRT(μ)	6	6	6	6	6	6

NN - หมายถึงการแจกแจงแบบปกติ

UU - หมายถึงการแจกแจงแบบยูนิฟอร์ม

LL - หมายถึงการแจกแจงแบบโลจิสติก

1.2 การทดสอบสแควร์แรงค์ (\bar{X}) สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีเมื่อขนาดของตัวอย่างเท่ากันทั้ง 2 กลุ่ม คือ 10 15 และ 20 ในกรณีที่ขนาดของตัวอย่างทั้งสองกลุ่มเป็น (5:10) (5:15) และ (5:20) การทดสอบสแควร์แรงค์ (\bar{X}) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้น้อยมาก โดยเฉพาะเมื่อใช้เกณฑ์ของ Cochran ซึ่งสรุปผลความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบสแควร์แรงค์ (\bar{X}) ได้ดังตาราง 5.2 สำหรับลักษณะที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ไม่ได้ ของการทดสอบสแควร์แรงค์ (\bar{X}) คือ ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองมากกว่าค่า α ที่กำหนด

ตาราง 5.2 แสดงจำนวนครั้งที่การทดสอบสแควร์แรงค์ (\bar{X}) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ จากการทดลองทั้งหมด 9 กรณี สำหรับแต่ละกลุ่มของขนาดของตัวอย่าง จำแนกตามระดับนัยสำคัญที่กำหนด

ระดับนัยสำคัญ (α)	ขนาดของตัวอย่าง	เกณฑ์ของ Bradley	เกณฑ์ของ Cochran
.01	(10:10) (15:15) (20:20)	9	9
	(5:10) (5:15) (5:20)	1	1
.05	(10:10) (15:15) (20:20)	9	7
	(5:10) (5:15) (5:20)	1	0
.10	(10:10) (15:15) (20:20)	9	8
	(5:10) (5:15) (5:20)	7	0

2. กรณีที่ประชากรมีการแจกแจงแบบเบ้ สรุปผลได้ดังนี้

2.1 การทดสอบเอฟสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีเมื่อความโด่งมีค่า 2.8 3.0 และ 3.2 และความเบ้มีค่า .25 .50 .75 ไม่ว่าขนาดของตัวอย่างเท่ากันหรือไม่เท่ากัน แต่ในกรณีที่ความเบ้มีค่า .25 .50 .75 และ 1.0 และความโด่งมีค่า 3.6 และ 4.2 การทดสอบเอฟสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้น้อยมาก โดยเฉพาะเมื่อใช้เกณฑ์ของ Cochran ซึ่งสรุปผลความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 แสดงได้ดังตาราง 5.3 สำหรับลักษณะที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ไม่ได้ของการทดสอบเอฟ มีทั้งกรณีที่ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

จากการทดลองมีค่ามากกว่าค่า α ที่กำหนด และน้อยกว่าค่า α ที่กำหนด เมื่อความเบ้มีค่า .25 .50 และ .75 และความโด่งมีค่า 2.8 3.0 และ 3.2 สำหรับกรณีที่ความเบ้มีค่า .25 .50 .75 และ 1.0 และความโด่งมีค่า 3.6 และ 4.2 สภาวะที่ควบคุมความคลาดเคลื่อนประเภทที่ 1 ไม่ได้ คือค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองมีค่ามากกว่าค่า α ที่กำหนด

ตาราง 5.3 แสดงจำนวนครั้งที่การทดสอบเอฟควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้จากการทดลองทั้งหมด 54 กรณี สำหรับความเบ้ (.25, .50, .75) และความโด่ง (2.8, 3.0, 3.2) และจากการทดลองทั้งหมด 48 กรณี สำหรับความเบ้ (.25, .50, .75, 1.0) และความโด่ง (3.6, 4.2) จำแนกตามระดับนัยสำคัญที่กำหนด

ระดับนัยสำคัญ	ความเบ้	ความโด่ง	เกณฑ์ของ Bradley	เกณฑ์ของ Cochran
.01	(.25, .50, .75)	(2.8, 3.0, 3.2)	45	40
	(.25, .50, .75, 1.0)	(3.6, 4.2)	6	6
.05	(.25, .50, .75)	(2.8, 3.0, 3.2)	54	42
	(.25, .50, .75, 1.0)	(3.6, 4.2)	16	2
.10	(.25, .50, .75)	(2.8, 3.0, 3.2)	54	44
	(.25, .50, .75, 1.0)	(3.6, 4.2)	28	1

2.2 การทดสอบสแควร์แรงค์ (\bar{X}) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี เมื่อขนาดของตัวอย่างเป็น (10:10) (15:15) และ (20:20) โดยที่ความเบ้มีค่า .25 และ .50 สำหรับทุก ๆ ค่าของความโด่งที่ใช้ในการวิจัยครั้งนี้ แต่เมื่อความเบ้มีค่า .75

และ 1.0 แม้ว่าขนาดของตัวอย่างจะเท่ากัน คือ (10:10) (15:15) และ (20:20) การทดสอบสแควร์แรงค์ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้น้อย สำหรับกรณีที่มีขนาดของตัวอย่างเป็น (5:10) (5:15) และ (5:20) การทดสอบสแควร์แรงค์ (X²) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้น้อย ทุก ๆ ค่าของความเบ้และความโด่งที่ใช้ในการวิจัยครั้งนี้ ซึ่งความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 แสดงได้ดังตาราง 5.4 และกรณีที่การทดสอบสแควร์แรงค์ (X²) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ไม่ได้ นั่นคือลักษณะที่ควบคุมไม่ได้คือค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 มากกว่าค่า α ที่กำหนด

ตาราง 5.4 แสดงจำนวนครั้งที่การทดสอบสแควร์แรงค์ (X²) ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ จากการทดลองทั้งหมด 30 กรณี สำหรับความเบ้ .25 และ .50 และ 21 กรณี สำหรับความเบ้ .75 และ 1.0 จำแนกตามระดับนัยสำคัญที่กำหนด

ระดับนัยสำคัญ	ขนาดของตัวอย่าง	ความเบ้	เกณฑ์ของ Bradley	เกณฑ์ของ Cochran
.01	(10:10), (15:15)	(.25, .50)	22	22
	(20:20)	(.75, 1.0)	3	3
	(5:10), (5:15)	(.25, .50)	7	7
	(5:20)	(.75, 1.0)	0	0
.05	(10:10), (15:15)	(.25, .50)	30	24
	(20:20)	(.75, 1.0)	7	1
	(5:10), (5:10)	(.25, .50)	6	0
	(5:20)	(.75, 1.0)	0	0
.10	(10:10), (15:15)	(.25, .50)	30	22
	(20:20)	(.75, 1.0)	8	0
	(5:10), (5:15)	(.25, .50)	22	0
	(5:20)	(.75, 1.0)	3	0

ผลสรุปอำนาจของการทดสอบ

จากการทดลองเพื่อหาค่าอำนาจของการทดสอบเอฟ การทดสอบคล้อยท้าย การทดสอบมัด การทดสอบทูกี้-ซีเกล และการทดสอบสแควร์แรงค์ เมื่อประชากรมีการแจกแจงแบบปกติ แบบยูนิฟอร์ม และแบบโลจิสติก และศึกษาอำนาจของการทดสอบเอฟและการทดสอบสแควร์แรงค์ (\bar{X}) เมื่อประชากรมีการแจกแจงแบบเบ้ สรุปผลได้ดังนี้

1. เมื่อประชากรมีการแจกแจงแบบปกติ การทดสอบเอฟเป็นการทดสอบที่มีอำนาจของการทดสอบสูงที่สุดเมื่อเทียบกับการทดสอบอื่น ๆ เมื่อขนาดของตัวอย่างเป็น (10:10) (15:15) และ (20:20) โดยมีอำนาจของการทดสอบสูงที่สุด 20 กรณี เมื่อ $\alpha = .01$ 21 กรณีเมื่อ $\alpha = .05$ และ $.10$ จากการทดลองหาค่าอำนาจของการทดสอบทั้งหมด 21 กรณี สำหรับกรณีที่ขนาดของตัวอย่างทั้งสองกลุ่มเป็น (5:10) (5:15) และ (5:20) การทดสอบสแควร์แรงค์ (μ) มีอำนาจของการทดสอบสูงที่สุดเป็นส่วนมาก เมื่ออัตราส่วนของความแปรปรวนของประชากร $\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$ มากกว่า 1 โดยมีอำนาจของการทดสอบสูงที่สุด 21 กรณีเมื่อ $\alpha = .01$ 18 กรณีเมื่อ $\alpha = .05$ และ $.10$ จากการทดลองหาค่าอำนาจของการทดสอบทั้งหมด 21 กรณี แต่ในกรณีที่อัตราส่วนของความแปรปรวนของประชากรน้อยกว่า 1 การทดสอบเอฟมีอำนาจของการทดสอบสูงที่สุด

2. เมื่อประชากรมีการแจกแจงแบบยูนิฟอร์ม การทดสอบสแควร์แรงค์ (μ) มีอำนาจของการทดสอบสูงที่สุดเป็นส่วนมาก โดยเฉพาะเมื่อขนาดของตัวอย่างเป็น (5:10) (5:15) และ (5:20) และอัตราส่วนของความแปรปรวนของประชากรมีค่ามากกว่า 1 โดยที่การทดสอบสแควร์แรงค์ (μ) มีอำนาจของการทดสอบสูงที่สุด 21 กรณี เมื่อ $\alpha = .01$ และ $\alpha = .05$ และ 20 กรณีเมื่อ $\alpha = .10$ จากการทดลองหาค่าอำนาจของการทดสอบทั้งหมด 21 กรณี แต่เมื่ออัตราส่วนของความแปรปรวนมีค่าน้อยกว่า 1 การทดสอบคล้อยท้ายก็เป็นการทดสอบที่มีอำนาจของการทดสอบสูงเช่นเดียวกับการทดสอบสแควร์แรงค์ (μ) กล่าวคือ การทดสอบสแควร์แรงค์ (μ) มีอำนาจของการทดสอบสูงที่สุด 18 กรณี เมื่อ $\alpha = .01$ 8 กรณีเมื่อ $\alpha = .05$ และ 12 กรณีเมื่อ $\alpha = .10$ ในขณะที่การทดสอบคล้อยท้ายมีอำนาจของการทดสอบสูงที่สุด 3 กรณีเมื่อ $\alpha = .01$ 13 กรณีเมื่อ $\alpha = .05$ และ 9 กรณีเมื่อ $\alpha = .10$ จากการทดลองทั้งหมด 21 กรณี สำหรับกรณีที่ขนาดของตัวอย่างทั้งสองกลุ่มเป็น (10:10) (15:15) และ (20:20) การทดสอบสแควร์แรงค์ (μ) และการทดสอบคล้อยท้ายเป็นการทดสอบที่มีอำนาจของการทดสอบสูงที่สุด กล่าวคือ การทดสอบสแควร์แรงค์ (μ) มีอำนาจของการทดสอบสูงที่สุด 17 กรณี เมื่อ $\alpha = .01$ 13 กรณีเมื่อ $\alpha = .05$ และ 8 กรณีเมื่อ

$\alpha = .10$ และการทดสอบคลัสเตอร์ที่มีอำนาจของการทดสอบที่สูงที่สุด 4 กรณี เมื่อ $\alpha = .01$ 31 กรณีเมื่อ $\alpha = .05$ และ 13 กรณีเมื่อ $\alpha = .10$ จากการทดลองหาค่าอำนาจการทดสอบทั้งหมด 21 กรณี

3. เมื่อประชากรมีการแจกแจงแบบโลจิสติก โดยส่วนมากแล้วการทดสอบสแควร์แรงค์ (M) มีอำนาจของการทดสอบที่สูงที่สุด โดยมีการทดสอบสแควร์แรงค์ (M) มีอำนาจของการทดสอบที่สูงที่สุด 21 กรณี เมื่อ $\alpha = .01$ และ $\alpha = .05$ และ 18 กรณีเมื่อ $\alpha = .10$ สำหรับของตัวอย่างที่เป็น $(10:10)$ $(15:15)$ และ $(20:20)$ และในกรณีที่ขนาดของตัวอย่างเป็น $(5:10)$ $(5:15)$ และ $(5:10)$ การทดสอบสแควร์แรงค์ (M) มีอำนาจของการทดสอบที่สูงที่สุด 21 กรณี เมื่อ $\alpha = .01$ $.05$ และ $.10$ และอัตราส่วนความแปรปรวนของประชากรมากกว่า 1 แต่เมื่ออัตราส่วนของความแปรปรวนของประชากรน้อยกว่า 1 การทดสอบสแควร์แรงค์ (M) มีอำนาจของการทดสอบที่สูงที่สุด 20 กรณี เมื่อ $\alpha = .01$ 19 กรณี เมื่อ $\alpha = .05$ และ 17 กรณีเมื่อ $\alpha = .10$ จากการทดลองหาค่าอำนาจของการทดสอบ 21 กรณี และในกรณีของการแจกแจงแบบโลจิสติก การทดสอบสแควร์แรงค์ (\bar{X}) ก็มีอำนาจของการทดสอบสูงเช่นเดียวกับการทดสอบสแควร์แรงค์ (M) เมื่อการทดสอบนี้ควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

4. เมื่อประชากรมีการแจกแจงแบบวี ซึ่งปริยายเทียบอำนาจของการทดสอบเอฟ และการทดสอบสแควร์แรงค์ (\bar{X}) เมื่อการทดสอบทั้งสองวิธีควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี คือความเบ้มีค่า $-.25$ และ $-.50$ และความโด่งมีค่า 2.8 3.0 และ 3.2 และขนาดของตัวอย่างที่เป็น $(10:10)$ $(15:15)$ และ $(20:20)$ สำหรับขนาดของตัวอย่างที่เป็น $(5:10)$ การทดสอบทั้งสองวิธีควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีเมื่อความเบ้เป็น $-.25$ เท่านั้น ซึ่งจากค่าอำนาจของการทดสอบทั้งสองวิธี ปริมาณว่าในกรณีที่ขนาดของตัวอย่างเป็น $(10:10)$ $(15:15)$ และ $(20:20)$ การทดสอบเอฟมีอำนาจของการทดสอบที่สูงกว่า การทดสอบสแควร์แรงค์ (\bar{X}) เมื่อความเบ้มีค่า $-.25$ แต่ในกรณีที่ความเบ้มีค่า $-.50$ เมื่ออัตราส่วนของความแปรปรวนของประชากรทั้งสองกลุ่มมีค่าเป็น 2 มีบางกรณีที่อำนาจของการทดสอบทั้งสองวิธีใกล้เคียงกัน แต่เมื่ออัตราส่วนของความแปรปรวนของประชากรทั้งสองกลุ่มมีค่าอื่น ๆ การทดสอบเอฟมีอำนาจของการทดสอบที่สูงกว่า และในกรณีที่ขนาดของตัวอย่างทั้งสองกลุ่มเป็น $(5:10)$ การทดสอบสแควร์แรงค์ (\bar{X}) มีอำนาจของการทดสอบที่สูงกว่าการทดสอบเอฟเมื่ออัตราส่วนของความแปรปรวนของประชากรทั้งสองกลุ่มมีค่ามากกว่า 1 แต่เมื่ออัตราส่วนของความแปรปรวนของประชากรมีค่าน้อยกว่า 1 การทดสอบเอฟมีอำนาจของการทดสอบที่สูงกว่า

5. ในกรณีของการทดสอบแบบนอนพาราเมตริก คือ การทดสอบคล้อยท้าย การทดสอบมัด การทดสอบทูกี้-ซีเกล และการทดสอบสแควร์แรงค์ (μ) เมื่อพิจารณาค่าอำนาจของการทดสอบทั้ง 4 วิธีดังกล่าวในกรณีที่ขนาดของตัวอย่างทั้งสองกลุ่มเป็น (5:10) (5:15) (5:20) (10:10) (15:15) และ (20:20) ปรากฏว่าเมื่อเปรียบเทียบค่า ARE กับค่าอำนาจของการทดสอบสองวิธีแล้ว ค่า ARE สามารถที่จะใช้เป็นเกณฑ์ในการพิจารณาเปรียบเทียบอำนาจของการทดสอบระหว่างการทดสอบคล้อยท้าย การทดสอบมัด และการทดสอบทูกี้-ซีเกล แต่ในการเปรียบเทียบอำนาจของการทดสอบสแควร์แรงค์ (μ) และการทดสอบอื่น ๆ ค่า ARE ไม่เป็นเกณฑ์ที่ดีที่จะใช้ในการพิจารณาเปรียบเทียบอำนาจของการทดสอบ

6. รูปแบบของการแจกแจงของประชากรมีผลต่อค่าของอำนาจของการทดสอบ เช่นเมื่อประชากรมีการแจกแจงแบบปกติ แบบยูนิฟอร์ม และแบบโลจิสติก กล่าวคือ ในกรณีที่ประชากรมีการแจกแจงแบบยูนิฟอร์ม ค่าอำนาจของการทดสอบจะสูงกว่ากรณีที่ประชากรมีการแจกแจงแบบอื่น ๆ สำหรับกรณีที่ประชากรมีการแจกแจงแบบเบ้ ซึ่งศึกษาอำนาจของการทดสอบเมื่อความเบ้มีค่า .25 และ .50 และความโด่งมีค่า 2.8 3.0 และ 3.2 ซึ่งค่าอำนาจของการทดสอบแตกต่างกันไม่มากสำหรับการทดสอบเอฟ แต่อำนาจของการทดสอบสแควร์แรงค์ (\bar{X}) จะมีค่าน้อยลงเมื่อความโด่งของประชากรมีค่าสูงขึ้น

การอภิปรายผล

จากผลการวิจัยครั้งนี้ เมื่อพิจารณาถึงความแกร่งของการทดสอบ เมื่อประชากรมีการแจกแจงที่มีความเบ้และความโด่งต่าง ๆ กัน ปรากฏว่าความโด่งของประชากรมีผลต่อความแกร่งของการทดสอบเอฟมากกว่าความเบ้ของประชากร ซึ่งผลการวิจัยนี้สอดคล้องกับงานวิจัยของ Pearson และ Please (1975:223-291) แต่อย่างไรก็ตาม Pearson และ Please ศึกษาเฉพาะกรณีที่ขนาดของตัวอย่างเท่ากัน คือ 10 และ 25 เท่านั้น แต่ในการวิจัยครั้งนี้ศึกษาในกรณีที่ขนาดของตัวอย่างไม่เท่ากันด้วย คือ ขนาดของตัวอย่างทั้งสองกลุ่มเป็น (5:10) (5:15) และ (5:20) ซึ่งจากผลการวิจัยก็ได้ผลสรุป เช่นเดียวกับกรณีที่ขนาดของตัวอย่างเท่ากัน คือ ความโด่งของประชากรมีผลต่อความแกร่งของการทดสอบเอฟ และในการวิจัยครั้งนี้พบว่า ความเบ้ และขนาดของตัวอย่างมีผลต่อความแกร่งของการทดสอบสแควร์แรงค์ (\bar{X}) มากกว่าความโด่งของประชากร กล่าวคือ เมื่อความเบ้ของประชากรมีค่า

มากกว่า .50 หรือขนาดของตัวอย่างไม่เท่ากัน คือ (5:10) (5:15) (5:20) การทดสอบ
สแควร์แรงค์ (\bar{X}) จะควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่ดี

นอกจากนี้ เมื่อพิจารณาค่าของอำนาจของการทดสอบกับรูปแบบของการแจกแจงของ
ประชากรเมื่อประชากรมีการแจกแจงแบบปกติ แบบยูนิฟอร์ม และแบบโลจิสติก ซึ่งมีความเบ้
เป็น 0 และความโด่ง 3.0 1.8 และ 4.2 ตามลำดับ ปรากฏว่าอำนาจของการทดสอบ
จะมีค่าสูงที่สุดเมื่อประชากรมีการแจกแจงแบบยูนิฟอร์ม ในขณะที่ประชากรมีการแจกแจงแบบ
โลจิสติก ค่าอำนาจของการทดสอบจะต่ำ เมื่อเทียบกับการแจกแจงแบบปกติและแบบยูนิฟอร์ม
ซึ่งแสดงว่าความโด่งของประชากรมีผลต่อค่าอำนาจของการทดสอบ และในกรณีที่ประชากร
มีการแจกแจงแบบเบ้ นั้น เมื่อพิจารณาค่าอำนาจของการทดสอบเอฟและการทดสอบสแควร์แรงค์
(\bar{X}) เมื่อความเบ้มีค่า .25 และ .50 และความโด่งมีค่า 2.8 3.0 และ 3.2 และขนาด
ของตัวอย่างเป็น (10:10) (15:15) และ (20:20) ปรากฏว่าอำนาจของการทดสอบ
สแควร์แรงค์ (\bar{X}) มีค่าน้อยลงเมื่อความโด่งมีค่ามากขึ้น สำหรับอำนาจของการทดสอบเอฟ
เมื่อประชากรมีความเบ้และความโด่งดังกล่าวอาจจะเห็นความแตกต่างไม่น้อยของค่าอำนาจ
การทดสอบได้น้อยเมื่อความโด่งมีค่ามากขึ้น ทั้งนี้อาจจะเป็นเพราะว่าความโด่งของประชากร
ที่ศึกษาอำนาจของการทดสอบมีค่าใกล้เคียงกันมาก

ข้อ 1 ข้อเสนอแนะ

ในการเลือกใช้สถิติทดสอบการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม
นั้น การทดสอบเอฟเป็นการทดสอบที่ควรเลือกใช้ เมื่อประชากรมีการแจกแจงแบบปกติ หรือ
รูปแบบของการแจกแจงเบ้หรือโด่งจากการแจกแจงแบบปกติไม่มากนัก (เช่น ความเบ้อยู่ช่วง
(0-75) และความโด่งมีค่าอยู่ในช่วง (2.8-3.2)) สำหรับกรณีที่ความเบ้หรือความโด่งของ
ประชากรมีค่าต่างจากความเบ้และความโด่งของการแจกแจงแบบปกติมาก ไม่ควรเลือกใช้
การทดสอบเอฟ ทั้งนี้เพราะว่าความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 จะมีค่า
สูง ซึ่งหมายความว่าผู้วิจัยเสี่ยงต่อการปฏิเสธสมมติฐาน H_0 เมื่อสมมติฐาน H_0 ถูก ใน
กรณีนี้ผู้วิจัยอาจจะพิจารณาการทดสอบอื่นที่แกร่งกว่าการทดสอบเอฟ เช่น การทดสอบแบบ
นอนพาราเมตริก แต่ในกรณีที่ไมต้องการใช้การทดสอบแบบนอนพาราเมตริก อาจจะเลือก
ใช้การทดสอบที่ Brown และ Forsythe (1974:364-366) เป็นผู้เสนอ ซึ่งการทดสอบ
ที่ Brown และ Forsythe เสนอนั้นเป็นการทดสอบที่ดัดแปลงมาจากการทดสอบของลีเวน

(Levene's test) โดยใช้ค่ามีเดียนของกลุ่มตัวอย่างแทนที่จะใช้ค่าเฉลี่ยของกลุ่มตัวอย่าง ซึ่งพิจารณาได้จากสถิติทดสอบแต่ละวิธีได้ดังนี้

การทดสอบของ Levene

$$W_o = \frac{\sum_i n_i (\bar{Z}_{i.} - \bar{Z}_{..})^2 / (g-1)}{\sum_i \sum_j (Z_{ij} - \bar{Z}_{i.})^2 / \sum_i (n_i - 1)}$$

โดยที่ $Z_{ij} = |x_{ij} - \bar{x}_i|$; $j=1, \dots, n_i$; $i=1, \dots, g$

$$\bar{Z}_{i.} = \sum Z_{ij} / n_i$$

$$\bar{Z}_{..} = \sum \sum Z_{ij} / \sum n_i$$

n_i = จำนวนค่าสังเกตของกลุ่มตัวอย่างที่ i

g = จำนวนกลุ่มตัวอย่าง

โดยที่ค่าวิกฤต W_o หาค่าได้จากตารางเอฟ เมื่อจำนวนชั้นแห่งความเป็นอิสระเป็น $g-1$ และ $\sum (n_i - 1)$

การทดสอบของ Brown และ Forsythe

เหมือนกรณีของ Levene แต่เปลี่ยนค่า $Z_{ij} = |x_{ij} - \bar{x}_i|$ เป็น $Z_{ij} = |x_{ij} - \tilde{x}_i|$ โดยที่ \tilde{x}_i เป็นค่ามีเดียนของกลุ่มตัวอย่างจากประชากรที่ i

การทดสอบของ Brown และ Forsythe แม้ว่าจะเป็นการทดสอบซึ่งแกร่ง แต่เมื่อเปรียบเทียบอำนาจของการทดสอบเอฟ หรือการทดสอบของ Levene ในกรณีที่การทดสอบดังกล่าวควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้จะมีอำนาจของการทดสอบต่ำกว่าการทดสอบทั้ง 2 วิธีนี้