ความน่าเชื่อถือของผลการวิเคราะห์ด้วยสถิติทดสอบวิลคอกซัน – แมนวิทนีย์ เมื่อคำนึงถึงข้อกำหนดเบื้องต้น จากโปรแกรมสำเร็จรูป SPSS และ MINITAB

Reliability of Wilcoxon-Mann-Whitney Test in Statistical Packages SPSS and MINITAB When Considering the Assumption

อุมาพร จันทศร

Umaporn Chantasorn

สาขาวิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

สถิติทดสอบวิลคอกซัน – แมนวิทนีย์ (WMW) สามารถใช้ทดสอบถึงความเท่ากันของ ค่ากลางจากประชากร 2 กลุ่มที่เป็นอิสระกัน มีข้อกำหนดเบื้องต้นที่สำคัญคือ ประชากรทั้งสองกลุ่มนั้นต้องมาจากการแจกแจงเดียวกัน แต่ผู้นำไปใช้มักละเลยถึงข้อกำหนดนี้ ดังนั้นวัตถุประสงค์ของงานวิจัยนี้ เพื่อหาผลสรุปว่าโปรแกรม SPSS และ MINITAB ได้คำนึงถึงข้อกำหนดดังกล่าวหรือไม่ ข้อมูลสำหรับการศึกษาครั้งนี้ได้โดยวิธีการจำลองด้วยวิธีมอนติคาร์โล เมื่อกำหนดให้ประชากรมีการแจกแจงเหมือนกันและต่างกัน และกำหนดให้มีค่ากลางเท่ากันและต่างกัน จากการทดลองสุ่มตัวอย่าง 1,000 ชุดจากแต่ละประชากรที่กำหนดให้ เมื่อกำหนดขนาดตัวอย่างจากสองประชากร, (n_1,n_2) ด้วยขนาด (10,10), (10,20) และ (10,30) และใช้ระดับนัยสำคัญในทดสอบคือ 0.05 และ 0.10 ทำการเปรียบเทียบค่าระดับนัยสำคัญที่แท้จริง (Empirical α) และอำนาจการทดสอบที่แท้จริง (Empirical power) ที่ได้จากสถิติ WMW จากโปรแกรม SPSS และโปรแกรม MINITAB การศึกษาพบว่าสถิติทดสอบ WMW จากโปรแกรมทั้ง 2 สามารถควบคุมค่าความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ตามเกณฑ์ของ Cochran และ Bradley ได้เฉพาะกรณีที่สองประชากรมีการแจกแจงเหมือนกัน ในเกือบทุกรูปแบบของการศึกษา แต่ในกรณีการแจกแจงต่างกันเกือบทั้งหมดของกรณีที่ศึกษาพบว่า สามารถควบคุมค่าความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้เพียงร้อยละ9-20 เท่านั้น ส่วนอำนาจการทดสอบที่แท้จริง พบว่าส่วนใหญ่มีค่าสูงเข้าใกล้ค่า 1 ทั้งในกรณีการแจกแจงเหมือนกันหรือต่างกัน โดยโปรแกรม SPSS ให้ค่ามากกว่าโปรแกรม MINITAB เล็กน้อย

คำสำคัญ: สถิติทดสอบวิลคอกซัน – แมนวิทนีย์ โปรแกรม SPSS โปรแกรม MINITAB

E-mail address : kcumapor@kmitl.ac.th โทรศัพท์ : 0819144879 โทรสาร : 023298426

Abstract

The Wilcoxon-Mann-Whitney test (WMW test) can be used to examine the equality of location parameter (mean or median) of 2 independent populations. However, there exist a pre - condition which is always overlooked by user that the two populations must have the same distributional form. The objective of this research is therefore to see whether the statistical package SPSS and MINTAB have taken this into account. The Monte Carlo method is applied to obtain data for this study. Various alternative conditions concerning the population distribution (same or different) and given value of mean (equal or unequal) were examined. For each situation, 1000 iterations were examined by applying a different sample size ,(n_1,n_2); (10,10), (10,20), (10,30) and at .05 and .10 level of significance. Comparisons of Empirical α and Empirical power of the test among the WMW test statistics from SPSS and MINITAB were done. It is found that only the case which the two populations has the same distribution, WMW test statistic from 2 sofewares have ability to control probability of Type I error in any situation. Except the case where distribution is different that WMW test statistic for almost situations can control the probability of Type I error only in range 9-20 percent. It is also found that Empirical power in all cases have value approaching 1 for the same or different distributions and in any cases, SPSS gives slightly higher value than MINITAB.

Keywords : Wilcoxon-Mann-Whitney test SPSS MINITAB.

1. บทนำ

กระบวนการเปรียบเทียบ 2 กรรมวิธีโดยพิจารณาจากค่ากลาง มักเกิดขึ้นเสมอในแทบทุกสาขาวิชาอาทิ การเปรียบเทียบผลการลดความดันโลหิตของผู้ป่วยความดันโลหิตสูงจากยาสมุนไพรไทยกับยาต่างประเทศ การเปรียบเทียบระยะทางที่วิ่งได้ของการใช้น้ำมันเบนซินกับน้ำมันแบบไบโอดีเซล หรือผลผลิตที่ได้จากการใช้ปุ๋ยเคมีหรือปุ๋ยชีวภาพ เป็นต้น กระบวนการที่จะให้คำตอบคือ กระบวนการทดสอบสมมติฐานทางสถิติ ด้วยการสุ่มตัวอย่างจากแต่ละกรรมวิธีมาอย่างเป็นอิสระกัน และคำนวณหาค่าความน่าจะเป็นที่จะเป็นไปตามสมมติฐานที่ตั้งเพื่อทดสอบว่าค่ากลางของ 2 กรรมวิธีไม่แตกต่างกัน แล้วพิจารณาว่ามีค่ามากเพียงพอที่จะสนับสนุนสมมติฐานหรือไม่ ค่าความน่าจะเป็นนี้จะได้จากการแจกแจงของสถิติที่ใช้ทดสอบ ภายใต้ข้อกำหนดเบื้องต้นบางประการ เช่น สถิติทดสอบ t (t – test) มีข้อกำหนดเบื้องต้นว่า ตัวอย่าง 2 ชุดนั้นต้องถูกสุ่มมาจากประชากรแบบปกติ 2 ชุด ที่เป็นอิสระกันและก่อนที่จะใช้ก็ต้องเลือกว่าจะใช้สถิติ t แบบค่าความแปรปรวนของ 2 ประชากรนั้นมีค่าเท่ากัน หรือไม่เท่ากัน [1] แม้ว่าในปัจจุบันมีงานวิจัยสนับสนุน t – test นี้ว่ามีคุณสมบัติแกร่ง (Robust) คือ ในกรณีที่ข้อกำหนดของความแปรปรวนที่เท่ากัน และ

ขนาดตัวอย่างมีขนาดใหญ่พอ รวมทั้งมีขนาดเท่ากันจาก 2 ประชากร ไม่เป็นจริง t–test ก็ยังคงมีประสิทธิภาพเหมือนเดิม แต่จะไม่มีคุณสมบัติแกร่งเมื่อขนาดตัวอย่างต่างกัน (คือจะให้ค่า ความน่าจะเป็นที่จะปฏิเสธสมมติฐานเบื้องต้นที่เป็นจริง;α และความน่าจะเป็นที่จะปฏิเสธสมมติฐานเบื้องต้นที่เป็นเท็จ; 1-βที่เปลี่ยนไป)

ในกรณีที่ข้อกำหนดเบื้องต้นของสถิติทดสอบ t ไม่เป็นจริง นักสถิติจะแนะนำให้ใช้สถิติทดสอบแบบไม่ใช้พารามิเตอร์ ซึ่งจะมีข้อกำหนดเบื้องต้นน้อยกว่าสถิติ t เช่น สถิติทดสอบ วิลคอกซัน – แมนวิทนีย์ (WMW) ที่ถูกนำไปใช้อย่างแพร่หลายในสาขาวิชาต่างๆมากมาย [2] เพราะสามารถใช้ได้ในกรณีที่ตัวอย่างมีขนาดเล็ก และข้อมูลมีมาตราวัดต่ำ คือมาตราวัดแบบเรียงลำดับ แต่มีผลงานวิจัยของ Zimmerman [3] ที่ได้ข้อสรุปว่า การทดสอบนี้ไม่เหมาะสมที่จะใช้แทนที่ t – test ในกรณีที่ขนาดตัวอย่างไม่เท่ากัน และมีความแปรปรวนของประชากรต่างกัน แม้ว่าสถิติทดสอบนี้จะมีข้อกำหนดเบื้องต้นเพียงว่า การแจกแจงต้องมีลักษณะต่อเนื่อง (ไม่จำเป็นต้องทราบการแจกแจง) แต่ต้องมีการแจกแจงที่เหมือนกันใน 2 ประชากร (Identical Distribution) ถ้าข้อกำหนดเบื้องต้นเหล่านี้เป็นจริง การทดสอบนี้จะมีคุณสมบัติไม่เอนเอียง (Unbiased test) และคงเส้นคงวา (Consistent test) ข้อกำหนดเกี่ยวกับการแจกแจงที่เหมือนกันนี้ มักถูกละเลยที่จะตรวจสอบก่อนที่จะใช้สถิติทดสอบ WMW ทั้งที่มีสถิติทดสอบแบบไม่ใช้พารามิเตอร์อีกแบบหนึ่ง คือ Fligner – Policello [4] ที่สามารถใช้แทนที่ได้ ในกรณีที่ข้อกำหนดเกี่ยวกับการแจกแจงที่เหมือนกัน หรือความแปรปรวนเท่ากัน ไม่เป็นจริง หรืออาจใช้การทดสอบแบบ Permutation สำหรับข้อมูลที่มีค่าตัวแปรมีลักษณะต่อเนื่อง

ในปัจจุบันโปรแกรมสำเร็จรูปถูกใช้เป็นเครื่องมือวิเคราะห์ข้อมูลแทนที่การคำนวณด้วยมือ (Calculation by Hand) แต่ขั้นตอนวิธี (Algorithm) ของแต่ละโปรแกรมจะต่างกัน เช่น การกำหนด Default (เช่นใช้ระดับนัยสำคัญ .05 ในการทดสอบสมมติฐาน หรือการปรับค่าซ้ำ (Ties)) นักวิจัยมักนิยมใช้โปรแกรมสำเร็จรูปทางสถิติ เช่น SPSS, MINITAB ในการวิเคราะห์ข้อมูล ซึ่งการวิเคราะห์ด้วยสถิติทดสอบ WMW จากทั้ง 2 โปรแกรม ไม่มีการตรวจสอบข้อกำหนดเบื้องต้น รวมทั้งถ้าผู้วิจัยตรวจสอบเองแล้วพบว่าข้อกำหนดดังกล่าวไม่เป็นจริง ก็ไม่มีสถิติทดสอบอื่นที่จะใช้แทนที่สถิติทดสอบ WMW เลย (คือไม่มีสถิติทดสอบ Fligner – Policello หรือสถิติทดสอบแบบ Permutation) และไม่ควรที่จะกลับไปใช้สถิติทดสอบ t เพราะข้อกำหนดเกี่ยวกับการแจกแจงปกติไม่เป็นจริง ดังนั้นงานวิจัยนี้ จึงจะศึกษาว่าโปรแกรม SPSS และ MINITAB ได้สร้าง Algorithm โดยให้ความสำคัญเกี่ยวกับข้อกำหนดเบื้องต้นของสถิติทดสอบ WMW หรือไม่ เพราะมีผลการวิเคราะห์ทางสถิติของโปรแกรมสำเร็จรูปต่าง ๆ ที่มีข้อผิดพลาด หรือทำให้เข้าใจผิด (Misleading) ในหลายประเด็น เช่น การศึกษาของ Bergmann et al. [5] เกี่ยวกับการปรับค่าซ้ำ (Correction for ties) การปรับค่าต่อเนื่อง (Correction for continuity) และการปรับค่าซ้ำและค่าต่อเนื่อง (Correction for continuity and ties) ของสถิติทดสอบ WMW หรืองานของ Berdhard et al. [6] ที่ให้ข้อสรุปว่า เฉพาะกรณี

ขนาดตัวอย่างเล็กและไม่มีค่าซ้ำเท่านั้น ค่า p (p – value) ที่ได้จากโปรแกรมสำเร็จรูปด้วยสถิติแบบไม่ใช้พารามิเตอร์ จึงจะเป็นค่าที่ถูกต้องแท้จริง (exact p – value)

ผลจากการศึกษาครั้งนี้ จะทำให้นักวิจัยที่ใช้สถิติทดสอบWMW จากโปรแกรม SPSS หรือ MINITAB ได้ทราบถึงข้อจำกัดในการใช้ และระมัดระวังในการใช้ และได้ผลสรุปที่แน่ใจว่าจะถูกต้องอย่างแท้จริง

2. วัตถุประสงค์ของการวิจัย

วัตถุประสงค์หลักคือ ศึกษาว่า โปรแกรม SPSS และ MINITAB ได้คำนึงถึงข้อกำหนดเบื้องต้นของสถิติทดสอบ WMW เพียงใด โดยมีรายละเอียดดังนี้

2.1 ศึกษาถึงความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ของสถิติทดสอบ WMW ในกรณีที่ประชากรทั้งสองมีค่ากลางไม่ต่างกัน ซึ่งจะแยกเป็นกรณีที่ประชากรทั้งสองมาจากการแจกแจงที่เหมือนกันและต่างกัน โดยศึกษาที่ระดับนัยสำคัญ 0.05 และ 0.10 และที่ขนาดตัวอย่างเท่ากัน หรือต่างกัน

2.2 ศึกษาถึงอำนาจการทดสอบของสถิติทดสอบ WMW ในกรณีที่ประชากรทั้งสองมีค่ากลางต่างกันที่มาจากประชากรที่มีการแจกแจงเหมือนกันหรือต่างกัน และเงื่อนไขเช่นเดียวกับข้อ 2.1

2.3 เปรียบเทียบผลการวิเคราะห์ด้วยสถิติทดสอบ WMW จากโปรแกรม SPSS และ MINITAB โดยเปรียบเทียบทั้งความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบ

3. วิธีดำเนินการวิจัย

ขั้นตอนในการวิจัยแบ่งออกเป็น 4 ขั้นตอน ดังนี้

3.1 สร้างการแจกแจงของ 2 ประชากรด้วยการจำลองโดยใช้เทคนิคมอนติคาร์โล ตามลักษณะที่กำหนดในสถานการณ์ต่าง ๆ ดังนี้

3.1.1 กำหนดให้ 2 ประชากรมีการแจกแจงที่เหมือนกัน และ ต่างกัน ดังนี้

Uniform, Normal, t, Exponential, Chi-Square ด้วยค่ากลางที่เท่ากันและต่างกัน

3.1.2 กำหนดขนาดตัวอย่าง(n_1,n_2) ด้วยขนาด (10,10), (10,20) และ(10,30)

3.1.3 จำนวนทำซ้ำ เท่ากับ 1,000 รอบในแต่ละสถานการณ์

3.2 คำนวณค่าสถิติทดสอบและค่าพี (p – value)

โดยนำข้อมูลในข้อ 3.1ไปวิเคราะห์ด้วยโปรแกรม SPSS เพื่อทดสอบ H_0 : ค่ากลางของ 2 ประชากรไม่แตกต่างกัน โดยใช้เมนู Nonparametric Test ด้วยสถิติทดสอบ WMW และนำข้อมูลชุดเดิมนั้นวิเคราะห์ด้วยโปรแกรม MINITAB เมนู Basic Statistics ด้วยสถิติทดสอบ WMW ในแต่ละสถานการณ์

3.3 คำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ที่แท้จริง (Empirical α)

เมื่อใช้ข้อมูลจาก 2 ประชากรที่มีค่ากลางไม่ต่างกัน ทดสอบด้วยสถิติทดสอบ WMW ด้วย

โปรแกรม SPSS และ MINITAB เป็นจำนวน 1,000 ชุดตัวอย่าง ในแต่ละสถานการณ์ หาค่าความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ที่แท้จริง (Empirical α) โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานเบื้องต้น (H_0 : ค่ากลางของ 2 ประชากรไม่แตกต่างกัน) เมื่อกำหนด α = 0.05 หรือ 0.10 คือนับว่ามีจำนวนกี่ครั้งที่ได้ค่าพีน้อยกว่า .05 หรือ .10 และนำมาหารด้วยค่า 1,000

และนำค่าที่ได้ไปเทียบกับเกณฑ์ของ Cochran [7] และ Bradley [8] ว่าสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้หรือไม่

เมื่อ เกณฑ์ของ Cochran เป็นดังนี้ ที่ระดับนัยสำคัญ .10 ถ้าค่าที่ได้มีค่าในช่วง (.08 - .12)
ที่ระดับนัยสำคัญ .05 ถ้าค่าที่ได้มีค่าในช่วง (.04 - .06)

ส่วน เกณฑ์ของ Bradley เป็นดังนี้ ที่ระดับนัยสำคัญ .10 ถ้าค่าที่ได้มีค่าในช่วง (.05 - .15)
ที่ระดับนัยสำคัญ .05 ถ้าค่าที่ได้มีค่าในช่วง (.025 - .075)

3.4 คำนวณค่าอำนาจการทดสอบที่แท้จริง (Empirical 1- β) เมื่อใช้ข้อมูลจาก 2 ประชากรที่มีค่ากลางต่างกัน ทดสอบด้วยสถิติทดสอบ WMW ด้วยโปรแกรม SPSS และ MINITAB เป็นจำนวน 1,000 ชุดตัวอย่าง ในแต่ละสถานการณ์ หาค่าอำนาจการทดสอบที่แท้จริง โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐานเบื้องต้น (H_0 : ค่ากลางของ 2 ประชากรไม่แตกต่างกัน) เมื่อกำหนด α = 0.05 หรือ 0.10 คือนับว่ามีจำนวนกี่ครั้งที่ได้ค่าพีน้อยกว่า 0.05 หรือ 0.10 และนำมาหารด้วยค่า 1,000

จากค่าที่ได้ เปรียบเทียบระหว่างค่าจากโปรแกรม SPSS และ MINITAB ว่าโปรแกรมใดให้ค่าที่มากกว่า ซึ่งหมายความว่าให้ค่าอำนาจการทดสอบที่สูงกว่า

4. ผลการวิจัย

สามารถสรุปผลได้รายละเอียด เมื่อแยกเป็นกรณีต่าง ๆ ดังนี้

4.1 กรณีประชากร 2 กลุ่มที่มีการแจกแจงเหมือนกัน และมีค่ากลางเท่ากัน ซึ่งจะได้ค่า **Empirical α** ดังรายละเอียดจากตารางที่ 1 ต่อไปนี้

ตารางที่ 1. ค่า Empirical α ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงเหมือนกัน และค่ากลางเท่ากัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	0.10	0.05	0.10
	$X_1 \sim U(0,5) + X_2 \sim U(0,5)$			
(10,10)	.044	.088	.033	.069
(10,20)	.064	.130	.057	.130
(10,30)	.042	.090	.041	.086

ตารางที่ 1. (ต่อ) ค่า Empirical α ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงเหมือนกัน และค่ากลางเท่ากัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
	X_1~ U(2,10) + X_2~ U(2,10)			
(10,10)	.053	.099	.047	.085
(10,20)	.045	.096	.041	.096
(10,30)	.048	.097	.047	.088
	X_1~ N(100,10) + X_2~ N(100,10)			
(10,10)	.055	.104	.045	.085
(10,20)	.035	.094	.032	.094
(10,30)	.064	.114	.062	.108
	X_1~ N(100,10) + X_2~ N(100,20)			
(10,10)	.066	.121	.055	.100
(10,20)	.036	.078	.031	.077
(10,30)	.019	.046	.013	.042
	X_1~ N(150,10) + X_2~ N(150,10)			
(10,10)	.058	.106	.043	.085
(10,20)	.049	.101	.046	.088
(10,30)	.047	.093	.050	.091
	X_1~ N(150,20) + X_2~ N(150,20)			
(10,10)	.052	.100	.041	.083
(10,20)	.049	.088	.042	.085
(10,30)	.052	.097	.045	.087
	X_1~ t(10) + X_2~ t(10)			
(10,10)	.049	.095	.042	.085
(10,20)	.044	.084	.041	.084
(10,30)	.047	.097	.046	.092
	X_1~ t(20) + X_2~ t(20)			

ตารางที่ 1. (ต่อ) ค่า Empirical α ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงเหมือนกัน และค่ากลางเท่ากัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
(10,10)	.052	.106	.046	.093
(10,20)	.052	.091	.043	.091
(10,30)	.045	.101	.042	.089
	X_1~ Exp(5) + X_2~ Exp(5)			
(10,10)	.059	.113	.046	.098
(10,20)	.043	.108	.041	.108
(10,30)	.041	.092	.039	.084
	X_1~ Exp(8) + X_2~ Exp(8)			
(10,10)	.062	.120	.057	.110
(10,20)	.045	.098	.042	.097
(10,30)	.064	.110	.059	.105
	X_1~ Chi(15) + X_2~ Chi(15)			
(10,10)	.052	.097	.044	.080
(10,20)	.064	.114	.055	.114
(10,30)	.043	.089	.036	.085
	X_1~ Chi(30) + X_2~ Chi(30)			
(10,10)	.056	.111	.048	.092
(10,20)	.053	.107	.049	.107
(10,30)	.051	.098	.047	.091
X_1~ t(10) + X_2~ t(20)				
(10,10)	.042	.088	.035	.077
(10,20)	.039	.083	.036	.083
(10,30)	.043	.094	.042	.089

จากตารางที่ 1 จะพบว่าผลสรุปจากโปรแกรม SPSS และ MINITAB ได้ผลคล้ายกันทุกการแจกแจง คือได้ค่า Empirical α ต่ำกว่าที่กำหนด คือ 0.05 และ 0.10 โดยโปรแกรม MINITAB จะให้ค่าต่ำกว่าโปรแกรม SPSS

ในเกือบทุกการแจกแจง และที่ทุกระดับนัยสำคัญ เช่นที่ประชากรแบบ U(0,5) และ U(0,5) ที่ $\alpha = 0.05$ และ (n_1,n_2) คือ (10,10) โปรแกรม SPSS จะได้ค่า 0.044 ในขณะที่ โปรแกรม MINITAB ได้ค่า 0.033 เป็นต้น เช่นเดียวกันที่ระดับนัยสำคัญ 0.10โปรแกรม SPSS ให้ค่า 0.088 แต่ โปรแกรม MINITAB ให้ค่า 0.069 มีบางกรณีคือ เพียง 9 กรณีจากทั้งหมด 78 กรณีเท่านั้นที่ โปรแกรม MINITAB ให้ค่าเท่ากับค่าจากโปรแกรม SPSS และมีเพียงกรณีเดียวเท่านั้นที่เท่านั้นที่ โปรแกรม MINITAB ให้ค่ามากกว่าค่าจากโปรแกรม SPSS

เมื่อใช้เกณฑ์ของ Cochran มาพิจารณาถึงความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 จะพบว่าโปรแกรม SPSS มี 30 ครั้งจาก 39 ครั้งที่ให้ค่า Empirical α ในขอบเขต 0.04 - 0.06 คิดเป็นร้อยละ 76.92 เมื่อใช้ระดับนัยสำคัญ 0.05 ส่วนที่ระดับความสำคัญ 0.10 มีถึง 35 ครั้ง ซึ่งคิดเป็นร้อยละ 89.74 ที่ให้ค่า Empirical α ในขอบเขต 0.08 - 0.12

ส่วนโปรแกรม MINITAB จะพบว่ามี 30 ครั้ง ที่ให้ค่า Empirical α ในขอบเขต 0.04 - 0.06 คิดเป็นร้อยละ 76.92 ที่ระดับนัยสำคัญ 0.05 ส่วนที่ระดับนัยสำคัญ 0.10 จะได้ค่าร้อยละของค่าที่อยู่ในขอบเขตเท่ากับ 87.18

เมื่อใช้เกณฑ์ของ Bradley โปรแกรมทั้งสองสามารถควบคุมค่าได้เกือบทั้งหมด ยกเว้นเพียงกรณีเดียวจากทั้งหมด 39 กรณี คิดเป็นร้อยละ97.43 ทั้งที่ระดับนัยสำคัญ 0.05 และ 0.10

เมื่อพิจารณาที่รูปแบบของ(n_1,n_2) คือเป็นรูปแบบ (10,10) หรือ (10,20) หรือ (10,30) พบว่าโปรแกรม MINITAB ให้จำนวนครั้งที่ได้ Empirical α ที่อยู่ในขอบเขตที่กำหนดตามเกณฑ์ของ Cochran เป็นจำนวน 22,21 และ 21 ครั้ง (จาก 26 ครั้งในแต่ละกรณี ของ (n_1,n_2)) ตามลำดับ ส่วนโปรแกรม SPSS ได้จำนวนครั้งเป็น 23,19 และ 22 ตามลำดับ

4.2 กรณีประชากร 2 กลุ่มมีการแจกแจงเหมือนกัน แต่มีค่ากลางต่างกัน ซึ่งจะได้ค่า Empirical power ดังรายละเอียดจากตารางที่ 2 ต่อไปนี้

ตารางที่ 2. ค่า Empirical power ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงเหมือนกัน แต่ค่ากลางต่างกัน

ขนาดตัวอย่าง (n_1,n_2)	**SPSS**		**MINITAB**	
	เมื่อกำหนดระดับนัยสำคัญ		**เมื่อกำหนดระดับนัยสำคัญ**	
	0.05	**0.10**	**0.05**	**0.10**
	$X_1 \sim U(0,5) + X_2 \sim U(2,10)$			
(10,10)	.951	.976	.944	.970
(10,20)	.992	.997	.990	.997
(10,30)	.997	1.000	.997	1.000
	$X_1 \sim N(100,10) + X_2 \sim N(150,10)$			

ตารางที่ 2. ค่า Empirical power ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงเหมือนกัน แต่ค่ากลางต่างกัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ N(100,10) + X_2~ N(150,20)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ Exp(5) + X_2~ Exp(8)			
(10,10)	.148	.228	.125	.203
(10,20)	.159	.241	.143	.241
(10,30)	.184	.279	.174	.271
	X_1~ Chi(15) + X_2~ Chi(30)			
(10,10)	.994	.988	.993	.997
(10,20)	.999	1.000	.999	1.000
(10,30)	1.000	1.000	1.000	1.000

จากตารางที่ 2 พบว่าค่า Empirical power ที่ได้จากโปรแกรม SPSS และ MINITAB ได้ค่าใกล้เคียงกัน คือ โดยส่วนใหญ่ได้ค่าเข้าใกล้ค่า 1 โดยโปรแกรม MINITAB ให้ค่าต่ำกว่าค่าจากโปรแกรมSPSS เล็กน้อย เช่น สุ่มตัวอย่างจากการแจกแจง U(0,5) และ U(2,10) ที่ระดับนัยสำคัญ 0.05 ที่ (n_1,n_2) = (10,10) ได้ค่า Empirical Power จากโปรแกรม SPSS = 0.951 ในขณะที่โปรแกรม MINITAB ให้ค่า 0.944 เป็นต้น และยังให้ละเอียดต่อไปว่า เมื่อเพิ่มระดับนัยสำคัญเป็น 0.10 และรูปแบบของ (n_1,n_2) มีความแตกต่างมากขึ้น คือเป็น(10,20) หรือ (10,30) จะได้ค่า Empirical Power สูงขึ้น ยกเว้นกรณีการแจกแจงปกติ จะให้อำนาจการทดสอบเท่ากัน คือมีค่าเป็น 1 หรือ 100% ทุกค่า α = 0.05 และ 0.10 และทุกแบบของ (n_1,n_2)

ยกเว้น ถ้าสุ่มตัวอย่างจาก Exp(5) + Exp(8) จะได้ค่า Empirical power ค่อนข้างต่ำ เช่น จากการแจกแจง Exp(5) + Exp(8) ที่ระดับนัยสำคัญ 0.05 และ (n_1,n_2) คือ (10,10) โปรแกรม SPSS ให้ค่า เท่ากับ 0.148 ในขณะที่ โปรแกรม MINITAB ให้ค่า 0.125 เป็นต้น ผลที่ได้แตกต่างจากการแจกแจงอื่นๆ ที่ศึกษาในครั้งนี้ ซึ่งจะต้องทำการศึกษาในรายละเอียดในครั้งต่อไป

4.3 กรณีประชากร 2 กลุ่มที่มีการแจกแจงต่างกันแต่มีค่ากลางเท่ากัน ซึ่งจะได้ค่า Empirical α ดังรายละเอียดจากตารางที่ 3 ต่อไปนี้

ตารางที่ 3. ค่า Empirical α ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงต่างกัน แต่ค่ากลางเท่ากัน

ขนาดตัวอย่าง(n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
	$X_1 \sim U(0,20) + X_2 \sim N(10,2)$			
(10,10)	.094	.146	.083	.139
(10,20)	.088	.184	.079	.162
(10,30)	.137	.360	.122	.214
	$X_1 \sim U(0,20) + X_2 \sim Exp(10)$			
(10,10)	.096	.312	.076	.134
(10,20)	.154	.389	.128	.156
(10,30)	.291	.172	.125	.071
	$X_1 \sim U(0,20) + X_2 \sim Chi(10)$			
(10,10)	.083	.265	.049	.175
(10,20)	.217	.411	.131	.215
(10,30)	.232	.284	.113	.198
	$X_1 \sim N(10,2) + X_2 \sim Exp(10)$			
(10,10)	.195	.264	.192	.262
(10,20)	.187	.299	.183	.291
(10,30)	.166	.256	.158	.249
	$X_1 \sim N(10,2) + X_2 \sim Chi(10)$			
(10,10)	.081	.138	.090	.144
(10,20)	.047	.110	.050	.120
(10,30)	.047	.095	.044	.111
	$X_1 \sim Exp(10) + X_2 \sim Chi(10)$			
(10,10)	0.120	0.210	0.122	0.215
(10,20)	0.174	0.259	0.179	0.262

ตารางที่ 3. ค่า Empirical α ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงต่างกัน แต่ค่ากลางเท่ากัน

ขนาดตัวอย่าง(n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
(10,30)	0.204	0.306	0.207	0.309
	$X_1 \sim N(0,2) + X_2 \sim t(10)$			
(10,10)	.085	.183	.067	.174
(10,20)	.112	.174	.081	.143
(10,30)	.148	.235	.116	.209

จากตารางที่ 3 จะพบว่า ทั้งสองโปรแกรมให้ค่า Empirical α ใกล้เคียงกัน และเกือบทั้งหมดของกรณีที่ศึกษาโปรแกรม SPSS จะให้ค่าใหญ่กว่าค่าจากโปรแกรม MINITABในทุกระดับนัยสำคัญ และทุกรูปแบบของ (n_1,n_2)

เมื่อใช้เกณฑ์ของ Cochran จะพบว่าเกือบทั้งหมดมีค่าสูงกว่าขอบเขตบน (คือค่า .06 หรือ .12 ที่ระดับนัยสำคัญ0.05 และ 0.10 ตามลำดับ) จะพบว่าโปรแกรม SPSS ให้ค่าที่อยู่ในขอบเขตของ Cochran เพียงร้อยละ 9.52 ทั้งที่ระดับนัยสำคัญ 0.05 และ 0.10 ส่วนโปรแกรม MINITAB ให้ค่าที่อยู่ในขอบเขตด้วยร้อยละ 14.28 และ 9.52 ที่ระดับนัยสำคัญ 0.05 และ 0.10 ตามลำดับ

ส่วนเกณฑ์ของ Bradley โปรแกรม SPSS ให้ค่าที่อยู่ในขอบเขต เพียงร้อยละ 9.52 และ 19.04 ที่ระดับนัยสำคัญ 0.05 และ 0.10 ตามลำดับ ส่วนโปรแกรม MINITAB ให้ค่าที่อยู่ในขอบเขตด้วยร้อยละ 19.04 และ33.33ที่ระดับนัยสำคัญ 0.05 และ 0.10 ตามลำดับ

4.4 กรณีประชากร 2 กลุ่มมีการแจกแจงต่างกัน และมีค่ากลางต่างกัน ซึ่งจะได้ค่า Empirical power ดังรายละเอียดจากตารางที่ 4 ต่อไปนี้

ตารางที่ 4. ค่า Empirical power ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงต่างกันและค่ากลางต่างกัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
$X_1 \sim U(0,10) + X_2 \sim N(10,2)$				
(10,10)	.983	.995	.978	.993
(10,20)	.998	1.000	.998	1.000
(10,30)	1.000	1.000	1.000	1.000
$X_1 \sim U(0,10) + X_2 \sim t(10)$				

ตารางที่ 4. ค่า Empirical power ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงต่างกันและค่ากลางต่างกัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
(10,10)	.998	1.000	.997	1.000
(10,20)	1.000	1.000	.999	1.000
(10,30)	1.000	1.000	1.000	1.000
$X_1 \sim U(0,10) + X_2 \sim Exp(10)$				
(10,10)	.170	.275	.149	.252
(10,20)	.202	.324	.185	.324
(10,30)	.983	.995	.982	.994
$X_1 \sim U(0,10) + X_2 \sim Chi(10)$				
(10,10)	.782	.883	.745	.850
(10,20)	1.000	1.000	1.000	1.000
(10,30)	.948	.979	.943	.976
	$X_1 \sim U(0,10) + X_2 \sim N(20,2)$			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	$X_1 \sim U(0,10) + X_2 \sim t(20)$			
(10,10)	.999	.999	.999	.999
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	$X_1 \sim U(0,10) + X_2 \sim Exp(20)$			
(10,10)	.651	.763	.621	.737
(10,20)	.778	.868	.762	.868
(10,30)	.865	.935	.859	.928
	$X_1 \sim U(0,10) + X_2 \sim Chi(20)$			
(10,10)	1.000	1.000	1.000	1.000

ตารางที่ 4. ค่า Empirical power ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงต่างกันและค่ากลางต่างกัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ N(10,2) + X_2~ t(20)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ N(10,2) + X_1~ Exp(20)			
(10,10)	.184	.265	.180	.261
(10,20)	.147	.249	.145	.249
(10,30)	.113	.220	.111	.220
	X_1~ N(10,2) + X_2~ Chi(20)			
(10,10)	.997	1.000	.997	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ t(10) + X_2~ Exp(20)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ t(10) + X_2~ Chi(20)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ Exp(10) + X_2~ Chi(20)			
(10,10)	0.820	0.886	0.825	0.890
(10,20)	0.888	0.928	0.891	0.931
(10,30)	0.906	0.956	0.910	0.961

ตารางที่ 4. (ต่อ) ค่า Empirical power ที่ได้จากการใช้สถิติทดสอบ WMW จาก 2 ประชากรที่มีการแจกแจงต่างกันและค่ากลางต่างกัน

ขนาดตัวอย่าง (n_1,n_2)	SPSS		MINITAB	
	เมื่อกำหนดระดับนัยสำคัญ		เมื่อกำหนดระดับนัยสำคัญ	
	0.05	**0.10**	**0.05**	**0.10**
	X_1~ U(0,20) + X_2~ t(10)			
(10,10)	.671	.942	.596	.784
(10,20)	.928	.991	.719	.952
(10,30)	.947	.985	.859	.934
	X_1~ N(10,2) + X_2~ t(10)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ t(10) + X_2~ Exp(10)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000
	X_1~ t(10) + X_2~ Chi(10)			
(10,10)	1.000	1.000	1.000	1.000
(10,20)	1.000	1.000	1.000	1.000
(10,30)	1.000	1.000	1.000	1.000

จากตารางที่ 4 ผลสรุปจากการใช้สถิติทดสอบ WMW จากโปรแกรม SPSS และ MINITAB ให้ผลคล้ายกันทุกกรณีของค่าระดับนัยสำคัญ และรูปแบบของ (n_1,n_2) คือมีค่า Empirical Power ค่อนข้างสูงมากเข้าใกล้ 1 โดยมีค่าเพิ่มขึ้น เมื่อกำหนดระดับนัยสำคัญให้สูงขึ้น จาก 0.05 เป็น 0.10 และความแตกต่างระหว่าง(n_1, n_2) มากขึ้น

5. สรุปผล

จากการศึกษาครั้งนี้ สามารถสรุปผลตามวัตถุประสงค์ของงานวิจัย ได้ดังนี้

5.1 ความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ของสถิติทดสอบวิลคอกซัน–แมนวิทนีย์ ในกรณีที่ประชากรสองกลุ่มมีการแจกแจงเหมือนกันได้ผลสรุปว่า สถิติทดสอบวิลคอกซัน–แมนวิทนีย์ มีความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้ค่อนข้างดี คือ

เมื่อพิจารณาตามเกณฑ์ของ Cochran โปรแกรม SPSS คำนวณค่า Empirical α ที่ควบคุมได้ตามเกณฑ์ด้วยร้อยละ 76.92 และ 89.74 เมื่อใช้ระดับนัยสำคัญในการทดสอบ 0.05 และ 0.10 ตามลำดับ ส่วนโปรแกรม MINITAB คำนวณค่าได้ร้อยละ 76.92 และ 87.18 ตามลำดับ

และเมื่อพิจารณาตามเกณฑ์ของ Bradley โปรแกรมทั้งสองคำนวณค่า Empirical α ที่ควบคุมได้เท่ากับร้อยละ 97.43 เท่ากัน ที่ระดับนัยสำคัญในการทดสอบ 0.05 และ 0.10

ส่วนรูปแบบของขนาดตัวอย่าง (n_1,n_2) ที่ใช้ต่างกัน คือเป็น (10,10) หรือ (10,20) หรือ (10,30) ที่ศึกษาในครั้งนี้ พบว่าโปรแกรมทั้งสองให้ค่า Empirical α ที่ควบคุมได้ตามเกณฑ์ของ Cochran ไม่ต่างกันมากนัก และเกณฑ์ของ Bradley พบว่าทุกรูปแบบของขนาดตัวอย่าง (n_1,n_2) สามารถควบคุมค่า Empirical α ได้เกือบจะเท่ากัน

แต่ในกรณีที่ประชากรสองกลุ่มมีการแจกแจงต่างกัน สถิติทดสอบวิลคอกซัน–แมนวิทนีย์ จะไม่สามารถควบคุมความคลาดเคลื่อนชนิดที่1ได้ โดยพบว่าเกือบทั้งหมดให้ค่าในขอบเขตตามเกณฑ์ของ Cochran หรือ Bradley ได้เพียงร้อยละ 9.52-33.33 เท่านั้น ซึ่งเกือบทั้งหมด มีค่าต่ำกว่าร้อยละ 20

5.2 ค่าอำนาจการทดสอบของสถิติทดสอบวิลคอกซัน–แมนวิทนีย์ จากโปรแกรม SPSS และ MINITAB พบว่า ในกรณีที่ประชากรมีการแจกแจงเหมือนกันและต่างกัน โปรแกรมทั้งสองให้ค่า Empirical Power ของสถิติทดสอบวิลคอกซัน–แมนวิทนีย์ ใกล้เคียงกัน คือส่วนใหญ่มีค่าเข้าใกล้ค่า1 และพบว่ามีค่าใหญ่ขึ้นเมื่อใช้ระดับนัยสำคัญเพิ่มขึ้นจาก 0.05 เป็น 0.10 และใช้ขนาดตัวอย่าง (n_1,n_2) ที่ต่างกันมากขึ้น คือใช้ (10,30) จะได้ค่า Empirical Power สูงที่สุด เมื่อเทียบกับการใช้ขนาดตัวอย่าง (10,10) หรือ (10,20)

5.3 เปรียบเทียบผลการวิเคราะห์ด้วยสถิติทดสอบวิลคอกซัน–แมนวิทนีย์ จากโปรแกรม SPSS และ MINITAB

ในเรื่องความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่1

ในกรณีที่สุ่มตัวอย่างจาก 2 ประชากรที่มีการแจกแจงเหมือนกัน และ มีค่ากลางเท่ากัน โปรแกรมทั้งสองมีความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่1 ได้ใกล้เคียงกัน โดยมีรายละเอียดที่พบว่า โปรแกรม MINITAB ให้ค่า Empirical α ที่ควบคุมได้เมื่อใช้ระดับนัยสำคัญในการทดสอบ 0.05 และ 0.10 ต่างกันพอสมควร คือร้อยละ 76.92 และ 87.18 ตามลำดับ เช่นเดียวกับโปรแกรม SPSS ที่ให้ผลที่ต่างกัน ด้วยค่าร้อยละ 76.92 และ 89.74 ตามลำดับ

ส่วนรูปแบบของขนาดตัวอย่าง (n_1,n_2) ที่ใช้ พบว่าไม่มีความแตกต่างกันในเรื่องความสามารถในการควบคุมความคลาดเคลื่อนชนิดที่1 คือจะสุ่มตัวอย่างแบบใดคือ (10,10) หรือ (10,20) หรือ(10,30) จะได้ผลไม่ต่างกันมากนัก

ในกรณีที่สุ่มตัวอย่างจาก 2 ประชากรที่มีการแจกแจงต่างกัน แต่ มีค่ากลางเท่ากัน พบว่าโปรแกรมทั้งสองให้ค่า Empirical α ใกล้เคียงกัน และเกือบทั้งหมดของกรณีที่ศึกษา โปรแกรมทั้งสองให้ค่าที่อยู่นอกขอบเขตตามเกณฑ์ของ Cochran หรือ Bradley

ในเรื่องค่าอำนาจการทดสอบ
โปรแกรมทั้งสองให้ค่า Empirical power ที่ใกล้เคียงกัน คือโดยส่วนใหญ่ได้ค่าเข้าใกล้ค่า 1 ทั้งในกรณีที่ประชากรทั้งสองมีการแจกแจงเหมือนกันและต่างกัน

จากข้อสรุปตามวัตถุประสงค์ของงานวิจัยนี้ สามารถให้ข้อสรุปเกี่ยวกับความน่าเชื่อถือของผลการวิเคราะห์ด้วยสถิติ ทดสอบวิลคอกซัน–แมนวิทนีย์เมื่อคำนึงถึงข้อกำหนดเบื้องต้น จากโปรแกรมสำเร็จรูป SPSS และ MINITAB ได้ว่าสถิติทดสอบวิลคอกซัน – แมนวิทนีย์ จะให้ผลที่น่าเชื่อถือ เฉพาะกรณีที่สุ่มตัวอย่างจาก 2 ประชากรที่มีการแจกแจงเหมือนกันเท่านั้น และโปรแกรมทั้งสองให้ผลใกล้เคียงกัน ดังนั้นผู้ใช้อาจเลือกใช้โปรแกรมใดก็ได้

6. อภิปรายผล

จากการศึกษาครั้งนี้ จะพบว่าสถิติทดสอบวิลคอกซัน – แมนวิทนีย์ มีประสิทธิภาพดีเฉพาะกรณีที่ประชากร 2 กลุ่มมีการแจกแจงเหมือนกันเท่านั้น นั่นคือให้ผลสรุปที่ถูกต้องในทางทฤษฎี คือ ให้ค่า Empirical α ต่ำกว่าที่กำหนดไว้ (หรือมีความน่าจะเป็นที่จะเกิด Type I error ต่ำกว่าที่กำหนด หรือมีความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ได้) และให้ค่าอำนาจการทดสอบที่สูง (คือค่า $(1-\beta)$ เข้าใกล้ 1) สรุปได้ว่า ในกรณีการแจกแจงเหมือนกัน สถิติทดสอบวิลคอกซัน – แมนวิทนีย์ จะสามารถควบคุม Type I error และ Type II error ให้มีค่าต่ำได้ในทุกกรณีของระดับนัยสำคัญ ซึ่งผลสรุปนี้สนับสนุนข้อกำหนดเบื้องต้นของสถิติ WMW ผู้นำสถิตินี้ไปใช้จึงต้องให้ความสำคัญแก่ข้อกำหนดนี้ก่อนนำไปใช้

แต่ในกรณีที่ประชากรสองกลุ่มมีการแจกแจงต่างกัน สถิติทดสอบวิลคอกซัน–แมนวิทนีย์ จะไม่สามารถควบคุม Type I error ได้เลย แต่ยังให้ค่าอำนาจการทดสอบที่สูง นั่นคือให้ผลสรุปถูกต้องเฉพาะกรณีที่ค่ากลางต่างกันเท่านั้น

ดังนั้นการจะใช้สถิติ WMW เพื่อตรวจสอบความเท่ากันของประชากร 2 กลุ่มที่เป็นอิสระกัน จึงต้องคำนึงถึงความเหมือนกันของการแจกแจงของประชากรทั้งสองเป็นสำคัญ ผลสรุปในระดับประชากรจึงจะเชื่อมั่นได้ว่าสามารถควบคุม Type I error ได้จริงและยังมีอำนาจการทดสอบสูง โดยผู้ใช้อาจทำการเปรียบเทียบรูปโค้งการแจกแจงจากตัวอย่าง 2 ชุดนั้นว่าคล้ายคลึงกันหรือไม่ เป็นลำดับแรกก่อน หรืออาจใช้สถิติทดสอบเกี่ยวกับความเหมือนกันของการแจกแจง เช่นสถิติ Komogorov-Smirnov สำหรับ 2 กลุ่ม ซึ่งสามารถใช้ได้จากโปรแกรมสำเร็จรูปทางสถิติทั่วๆ ไป ถ้าพบว่าข้อกำหนดนี้ไม่เป็นจริง ผู้ใช้อาจจะใช้สถิติทดสอบแบบไม่ใช้พารามิเตอร์อื่นๆ เช่น Fligner – Policello หรือ สถิติทดสอบแบบ Permutation หรือสถิติทดสอบแบบ Welch ซึ่งต้องใช้จากโปรแกรมสำเร็จรูปทางสถิติเฉพาะทาง คือสถิติที่ไม่ใช้พารามิเตอร์

เมื่อทำการเปรียบเทียบกับผลการวิจัยของ Zimmerman [3]ซึ่งได้ข้อสรุปว่า สถิติ WMW นี้ไม่เหมาะสมที่จะนำไปใช้แทนที่การทดสอบแบบที (t – test) ในกรณีขนาดตัวอย่างไม่เท่ากัน และมีความแปรปรวนของประชากรต่างกัน พบว่าการวิจัยครั้งนี้สนับสนุนข้อพบของ Zimmerman คือ จากการศึกษาครั้งนี้ เมื่อกำหนดให้ประชากรทั้งสองมีการแจกแจงต่างกัน รวมทั้งมีค่าความแปรปรวนต่างกัน แต่มีค่ากลางเท่ากัน จะพบว่าทุกกรณีไม่สามารถควบคุมความคลาดเคลื่อนชนิดที่ 1 ได้เลย แต่ในประเด็นขนาดตัวอย่างต่างกัน (คือ กรณี $n_1 – n_2$ คือ 10 – 20 และ 10 – 30) การศึกษาครั้งนี้พบว่าไม่ต่างกันในความสามารถควบคุมความคลาดเคลื่อนชนิดที่1 คือจะสุ่มตัวอย่างแบบใดก็ได้ผลคล้ายกัน

7. ข้อเสนอแนะ

จากข้อกำหนดของสถิติทดสอบ WMW ที่กล่าวว่า 2 ประชากรนั้นต้องมีรูปร่างเหมือนกัน ดังนั้นในการวิจัยครั้งต่อไป จึงควรทดลองใช้ข้อมูลจาก 2 ประชากรที่มีค่าความแปรปรวนเท่ากัน (ซึ่งหมายถึงมีรูปร่างเหมือนกัน) เนื่องจากการศึกษาครั้งนี้เกือบทั้งหมดจะเป็นกรณีค่าความแปรปรวนต่างกัน รวมทั้งกรณีใช้ข้อมูลจาก 2 ประชากรทีมีการแจกแจงต่างกัน แต่ใช้ตัวอย่างขนาดใหญ่ ซึ่งจะทำให้ได้รายละเอียดของงานวิจัยมากขึ้น

เอกสารอ้างอิง

[1] Sprent , P., 1993. Applied Nonparametric Statistical Methods. 2nd ed. Chapman & Hall.

[2] Deshpande, J.V. , Gore, A.P. and Shanubhogue, A., 1995. Statistical Analysis of Nonnormal Data. New Age International Publishers Limited Wiley Eastern Limited.

[3] Zimmerman, D.W., 1992**.** Failure of the Man-Whitney Test: A note on the simulation study of Gibbons and Chakraborti . *Journal of Experimental Education* , 60(4), 359 – 364.

[4] Hollander, M.W. and Wolfe, D.A., 1999. Nonparametric Statistical Methods. 2nd ed. John Willey & Son, Inc.

[5] Bergmann, R., Ludbrook, J. and Spooren**,** W. P. J.M., 2000**.** Different outcomes of the Wilcoxon – Mann – Whitney Test from different packages. *Journal of the American Statistician*, 54(1), 72-77.

[6] Bernhard, G., Alle, M., Herbold, M.and Meyers, W.,1988. Investigation on the reliability of some elementary nonparametric methods in statistical analysis systems. *Statistical Software Newsletter*, 14, 19–26.

[7] Cochran,W.G.,1945. Some methods for strengthening the common χ^2 test . *Biometrics*, 10, 417-451.

[8] Bradley, J.V.,1978. Robustness?**.** *British Journal of Mathematical and Statistical Psychology*, 31, 321-339.