

Research Article

การเปรียบเทียบประสิทธิภาพของการทดสอบทีและการทดสอบ ผลบวกลำดับที่ของวิลค็อกซันสำหรับประชากรสองกลุ่มที่เป็น อิสระกันเมื่อข้อมูลเป็นจำนวนนับ

A comparing on performance of t-test and Wilcoxon rank sum test for two independent populations using counting data

เขมิกา อูระวงศ์¹, จุฬารัตน์ ชุมนวล^{1,2*}, จิรนนท์ เผ่าจรูญ¹, อัสลยา สิงห์บำรุง¹

Khemika Urawong¹, Jularat Chumnuat^{1,2*}, Chiranan Phaochamrun¹, Asalaya Singhabumrung¹

¹สาขาวิทยาศาสตร์การคำนวณ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ หาดใหญ่สงขลา 90112 ประเทศไทย

¹Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla 90112, Thailand

²หน่วยวิจัยสถิติและการประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ หาดใหญ่สงขลา 90112 ประเทศไทย

²Statistics and Applications Research Unit, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla 90112, Thailand

*E-mail: jularat.c@psu.ac.th

Received: 04/10/2021; Revised: 27/03/2022; Accepted: 06/05/2022

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรสองกลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ โดยการแจกแจงที่นำมาศึกษา คือ การแจกแจงทวินาม และการแจกแจงปัวซอง การเปรียบเทียบประสิทธิภาพของการทดสอบพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ ผลการศึกษาพบว่า การทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณีการศึกษา สำหรับกำลังการทดสอบพบว่า ในกรณีที่ขนาดตัวอย่างทั้งสองกลุ่มเท่ากัน กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินามและการแจกแจงปัวซองมีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างทั้งสองกลุ่มและขนาดอิทธิพลเพิ่มขึ้น และเมื่อพิจารณาอิทธิพลของอัตราส่วนของขนาด

ตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ที่มีต่อการทดสอบภายใต้ขนาดตัวอย่างรวมที่เท่ากันพบว่า การทดสอบของ การทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันจะมีค่าสูงสุดเมื่ออัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ($n_1 : n_2$) เท่ากับ 1 : 1 ยกเว้นกรณีที่ n_1 และ n_2 เท่ากับ 10 และจะมีค่าน้อยที่สุดเมื่ออัตราส่วนของขนาด ตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ($n_1 : n_2$) เท่ากับ 1 : 5 (หรือ 5 : 1) ทั้งในกรณีที่ข้อมูลมีการแจกแจงทวินามและกรณีที่ ข้อมูลมีการแจกแจงปัวซอง นอกจากนี้ยังพบว่า การทดสอบทีซึ่งเป็นการทดสอบแบบอิงพารามิเตอร์ให้ค่าการ ทดสอบสูงกว่าการทดสอบผลบวกลำดับที่ของวิลค็อกซันทุกกรณี

คำสำคัญ: การทดสอบที, การทดสอบผลบวกลำดับที่ของวิลค็อกซัน, ค่าการทดสอบ, ความผิดพลาดแบบที่ 1

Abstract

This research aimed to compare the performance of the t-test and Wilcoxon rank sum test for testing the difference between the mean values of two independent populations using counting data. The distribution considered in this study were binomial distribution and Poisson distribution. The performance of these two tests was compared considering the ability to control the probability of type I error and power. The results showed that the t-test and Wilcoxon rank sum test could control the probability of type I error for all situations. In the case of power, when sample sizes of two groups were equal, the powers of t-test and Wilcoxon rank sum test for testing the difference between the mean values of two independent populations using counting data tended to increase when sample sizes of two groups, effect size, and the value of parameter n increased. Considering the effect of the ratios of sample sizes between groups toward powers, the results showed that the powers of the t-test and Wilcoxon rank sum test became maximum when the ratio of sample sizes between groups ($n_1 : n_2$) was 1 : 1 except for n_1 and n_2 equal 10 and became minimum when the ratio of sample sizes between groups ($n_1 : n_2$) was 1 : 5 (or 5 : 1) for both binomial and Poisson data. Moreover, the t-test which is a parametric test yielded higher powers than the Wilcoxon rank sum test for all cases.

Keywords: t-test, Wilcoxon rank sum test, power of a test, type I error

บทนำ

วิธีทางสถิติแบบอิงพารามิเตอร์ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่มที่เป็นอิสระกันที่นิยมใช้กันอย่างแพร่หลาย คือ การทดสอบที (Two independent samples t-test) ซึ่งการทดสอบดังกล่าวมี ข้อตกลงเบื้องต้นที่สำคัญ คือ ข้อมูลตัวอย่างทั้งสองกลุ่มต้องมาจากประชากรที่มีการแจกแจงปกติ แต่ในทางปฏิบัติ ข้อมูลไม่ได้มีการแจกแจงปกติเสมอไปโดยเฉพาะข้อมูลในงานวิจัยเชิงประยุกต์ เช่น ข้อมูลจำนวนผู้เข้ารับบริการ

จำนวนสินค้าที่ชำรุด จำนวนผู้ป่วยติดเชื้อ COVID-19 รายใหม่ จำนวนครั้งการเข้ารับการรักษาในโรงพยาบาลในระยะเวลา 1 ปี จำนวนครั้งการพยายามฆ่าตัวตายของผู้ป่วยที่เป็นโรคซึมเศร้า เป็นต้น การใช้การทดสอบที่ซึ่งเป็นการทดสอบอิงพารามิเตอร์ (parametric test) ในการวิเคราะห์ข้อมูลดังกล่าวอาจส่งผลให้ข้อสรุปไม่ถูกต้อง ผลการวิเคราะห์ข้อมูลที่ได้ไม่น่าเชื่อถือและไม่สามารถนำไปอ้างอิงหรือไปอธิบายสิ่งที่ต้องการศึกษาได้ ทั้งนี้ เนื่องจากสถิติทดสอบที่ถูกสร้างขึ้นภายใต้เงื่อนไขที่ว่าข้อมูลต้องเป็นข้อมูลเชิงปริมาณแบบต่อเนื่องและมาจากประชากรที่มีการแจกแจงปกติที่ไม่ทราบค่าความแปรปรวนของประชากร ตัวสถิติทดสอบจึงจะมีการแจกแจงที่ (Student's t-distribution) ตามทฤษฎีทางสถิติและให้กำลังการทดสอบ (power of a test) สูง การนำการทดสอบที่มาใช้กับข้อมูลจำนวนนับซึ่งเป็นข้อมูลเชิงปริมาณแบบไม่ต่อเนื่องและไม่ได้มีการแจกแจงปกติ จึงไม่สามารถการันตีได้ว่า การทดสอบที่จะยังให้ประสิทธิภาพดีหรือให้กำลังการทดสอบที่สูงอย่างที่ควรจะเป็น ในบางชุดข้อมูล การทดสอบที่อาจยังให้กำลังการทดสอบสูง แต่ในบางชุดข้อมูล การทดสอบที่อาจให้กำลังการทดสอบต่ำ ดังนั้น เพื่อแก้ไขปัญหาดังกล่าว การทดสอบไม่อิงพารามิเตอร์ (non-parametric test) จึงเป็นอีกทางเลือกหนึ่งที่ถูกนำมาใช้ในกรณีที่ข้อมูลไม่เป็นไปตามเงื่อนไขหรือข้อตกลงเบื้องต้นของการทดสอบอิงพารามิเตอร์

สำหรับการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร 2 กลุ่มที่เป็นอิสระกัน ในกรณีที่ข้อมูลไม่ผ่านข้อตกลงเบื้องต้นของการแจกแจงปกติ การทดสอบผลบวกลำดับที่ของวิลค็อกซัน (Wilcoxon rank sum test) มักจะถูกนำมาใช้แทนการทดสอบที (t-test) โดยมีงานวิจัยหลาย ๆ งานที่แสดงให้เห็นว่า การทดสอบผลบวกลำดับที่ของวิลค็อกซันให้กำลังการทดสอบ (power of the test) ที่สูงกว่าการทดสอบทีในหลาย ๆ สถานการณ์ ยกตัวอย่างเช่น Blair & Higgins (1980) และ Bridge & Sawilowsky (1999) ได้ศึกษากำลังการทดสอบของการทดสอบที (t-test) และการทดสอบผลบวกลำดับที่ของวิลค็อกซัน (Wilcoxon rank sum test) ในกรณีที่ข้อมูลมีขนาดเล็กและในกรณีที่ข้อมูลเป็นข้อมูลเชิงปริมาณแบบต่อเนื่องที่ไม่ได้มีการแจกแจงปกติ ซึ่งผลการวิจัยพบว่า การทดสอบผลบวกลำดับที่ของวิลค็อกซันมีกำลังการทดสอบสูงกว่าการทดสอบทีในกรณีที่ข้อมูลมีขนาดเล็กและมีความแปรปรวนสูง Sangthong (2013) ได้ศึกษาประสิทธิภาพของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่มในกรณีที่ข้อมูลมีการแจกแจงปกติ ข้อมูลมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ และข้อมูลมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ ซึ่งผลการวิจัยพบว่า การทดสอบทีและการทดสอบ Van der Waerden มีกำลังการทดสอบสูงสุดและสามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อตัวอย่างมีขนาดเล็ก สำหรับทุกการแจกแจง ส่วนในกรณีที่ข้อมูลมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติและข้อมูลมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ การทดสอบแมนน์-วิทนียู (Mann-Whitney U test) มีกำลังการทดสอบสูงสุดและสามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อตัวอย่างมีขนาดกลางและขนาดใหญ่ นอกจากนี้ Sangthong (2014) ยังได้ศึกษาความแกร่งและกำลังการทดสอบของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากรสองกลุ่มสำหรับข้อมูลแบบลิเคิร์ท 5 ระดับ ซึ่งผลการวิจัยพบว่า การทดสอบทีมีกำลังการทดสอบสูงสุดและสามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อตัวอย่างมีขนาดเล็กสำหรับทุกการ

แจกแจง ส่วนการทดสอบแมนน์-วิตนีย์มีกำลังการทดสอบสูงสุดและสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ เมื่อข้อมูลมีการแจกแจงปกติและตัวอย่างมีขนาดใหญ่ และเมื่อข้อมูลมีการแจกแจงเบ้ซ้ายและความโด่งต่ำกว่า ปกติและตัวอย่างมีขนาดเล็ก Araveeporn et al. (2017) ได้ศึกษาประสิทธิภาพของการทดสอบซี (z-test) การทดสอบที (t-test) การทดสอบแบบสุ่ม (randomization test) และการทดสอบแมนน์-วิตนีย์ (Mann-Whitney U test) สำหรับทดสอบค่าเฉลี่ยหรือค่ากลางของประชากร 2 กลุ่มที่เป็นอิสระกัน โดยผลการวิจัยพบว่า การทดสอบซี การทดสอบที และการทดสอบแมนน์-วิตนีย์สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทั้งในกรณี ตัวอย่างขนาดเล็กและตัวอย่างขนาดใหญ่ และเมื่อข้อมูลทั้ง 2 กลุ่มมีความแปรปรวนไม่เท่ากัน การทดสอบซีมีกำลังการทดสอบสูงที่สุด และ Kim & Park (2019) ได้ทำการศึกษาเกี่ยวกับข้อดคลงเบื้องต้นของการทดสอบทีสำหรับ ประชากร 2 กลุ่มที่เป็นอิสระกันและพบว่า ถ้าอัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ($n_1 : n_2$) เท่ากับ 1 : 1 จะให้กำลังการทดสอบที่สูงที่สุด

ดังนั้น ในการวิจัยนี้จึงต้องการศึกษาประสิทธิภาพของการทดสอบทีและการทดสอบผลบวกลำดับที่ของ วิลค็อกชันสำหรับข้อมูลเชิงปริมาณไม่ต่อเนื่องที่ได้จากการนับ (counting data) โดยการแจกแจงที่นำมาพิจารณา คือ การแจกแจงทวินาม (Binomial distribution) และการแจกแจงปัวซอง (Poisson distribution) ซึ่งเป็นการแจกแจงความ น่าจะเป็นของตัวแปรสุ่มแบบไม่ต่อเนื่องที่สำคัญที่มีค่าของตัวแปรสุ่มเป็นจำนวนนับและเป็นการแจกแจงที่ใช้กัน อย่างแพร่หลายในงานวิจัยเชิงประยุกต์ ทั้งนี้ เพื่อเป็นแนวทางให้กับผู้วิเคราะห์ข้อมูลในการเลือกใช้สถิติทดสอบ สำหรับการทดสอบสมมุติฐานความแตกต่างระหว่างค่ากลางของประชากรสองกลุ่มที่เป็นอิสระกันสำหรับข้อมูล จำนวนนับได้อย่างเหมาะสม

วิธีการวิจัย

การวิจัยครั้งนี้ใช้การจำลองข้อมูล (simulation study) ในการตรวจสอบประสิทธิภาพของการทดสอบที และการทดสอบผลบวกลำดับที่ของวิลค็อกชันสำหรับประชากรสองกลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ โดยผู้วิจัยได้กำหนดขอบเขตและขั้นตอนการดำเนินการวิจัยดังนี้

ขอบเขตการวิจัย

1. ข้อมูลที่ศึกษามีการแจกแจงทวินาม (Binomial distribution) และการแจกแจงปัวซอง (Poisson distribution) โดยมีพารามิเตอร์ของการแจกแจงดังนี้

การแจกแจงทวินาม $(n, p_1, p_2) = (10, 0.1, 0.1), (10, 0.5, 0.5), (10, 0.7, 0.7), (50, 0.1, 0.1), (50, 0.5, 0.5), (50, 0.7, 0.7), (100, 0.1, 0.1), (100, 0.5, 0.5), (100, 0.7, 0.7), (10, 0.1, 0.125), (10, 0.5, 0.525), (10, 0.7, 0.725), (50, 0.1, 0.125), (50, 0.5, 0.525), (50, 0.7, 0.725), (100, 0.1, 0.125), (100, 0.5, 0.525), (100, 0.7, 0.725)$

การแจกแจงปัวซอง $(\lambda_1, \lambda_2) = (1, 1), (5, 5), (10, 10), (1, 1.25), (1, 1.5), (1, 2), (5, 8)$

2. ขนาดตัวอย่างที่ศึกษาจำแนกเป็น 2 กรณี คือ กรณีขนาดตัวอย่างเท่ากัน และกรณีขนาดตัวอย่างไม่เท่ากัน

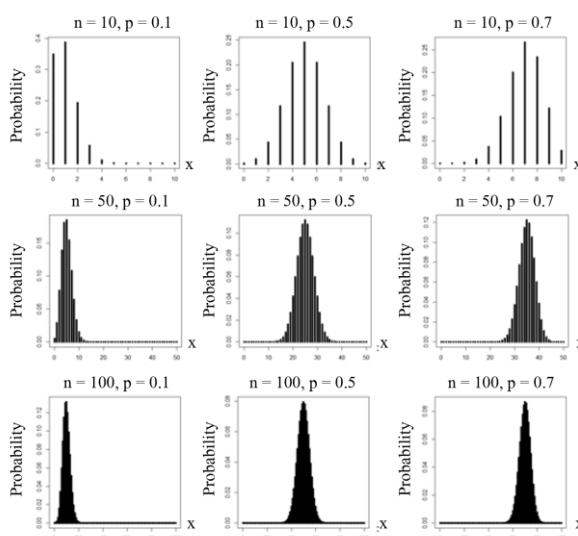
- 2.1 กรณีขนาดตัวอย่างเท่ากัน $(n_1, n_2) = (10, 10), (60, 60), (100, 100)$
- 2.2 กรณีขนาดตัวอย่างไม่เท่ากัน $(n_1, n_2) = (20, 100), (30, 90), (40, 80), (100, 20), (90, 30), (80, 40)$
3. การทดสอบที่ศึกษาเป็นการทดสอบแบบสองทาง (two-tailed test) โดยกำหนดระดับนัยสำคัญ $\alpha = 0.05$
4. ประสิทธิภาพของการทดสอบพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 (type I error) และกำลังการทดสอบ (power of a test)
5. ทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์ที่กำหนด

ขั้นตอนการดำเนินการวิจัย

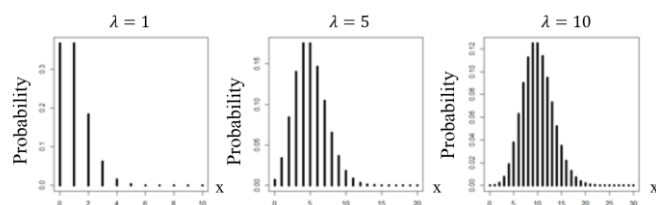
1. สร้างข้อมูลสำหรับใช้ในการวิจัยตามสถานการณ์ต่าง ๆ ที่กำหนดไว้ในขอบเขตการวิจัย กล่าวคือ สำหรับการแจกแจงทวินาม ทำการสร้างตัวแปรสุ่มตัวที่ 1 ที่มีการแจกแจงทวินามด้วยพารามิเตอร์ n และ p_1 และสร้างตัวแปรสุ่มตัวที่ 2 ที่มีการแจกแจงทวินามด้วยพารามิเตอร์ n และ p_2 โดยมีขนาดตัวอย่างเท่ากับ n_1 และ n_2 ตามลำดับ สำหรับการแจกแจงปัวซอง ทำการสร้างตัวแปรสุ่มตัวที่ 1 ที่มีการแจกแจงปัวซองด้วยพารามิเตอร์ λ_1 และสร้างตัวแปรสุ่มตัวที่ 2 ที่มีการแจกแจงปัวซองด้วยพารามิเตอร์ λ_2 โดยมีขนาดตัวอย่างเท่ากับ n_1 และ n_2 ตามลำดับ โดยในการสร้างตัวแปรสุ่มทั้ง 2 ตัว กำหนดให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรสุ่มทั้ง 2 ตัวเท่ากับ 0 เพื่อให้ตัวแปรสุ่มทั้ง 2 ตัว เป็นอิสระกัน และในการศึกษาความน่าจะเป็นของความผิดพลาดแบบที่ 1 ทำการสร้างข้อมูลภายใต้สมมติฐานว่าง (null hypothesis, H_0) ส่วนการศึกษาการทดสอบ ทำการสร้างข้อมูลภายใต้สมมติฐานทางเลือก (alternative hypothesis, H_1) โดยสมมติฐานของการทดสอบ คือ

H_0 : ประชากรทั้งสองกลุ่มมีค่ากลาง (ค่าเฉลี่ย/มัธยฐาน) ไม่แตกต่างกัน

H_1 : ประชากรทั้งสองกลุ่มมีค่ากลาง (ค่าเฉลี่ย/มัธยฐาน) แตกต่างกัน



รูปที่ 1 ข้อมูลที่มีการแจกแจงทวินามภายใต้ขอบเขตการวิจัย



รูปที่ 2 ข้อมูลที่มีการแจกแจงปัวซองภายใต้ขอบเขตการวิจัย

2. จำนวนค่าสถิติทดสอบที่และสถิติทดสอบผลบวกลำดับที่ของวิลค็อกซัน

2.1 สถิติทดสอบที่ (Moore, McCabe & Craig, 2014)

2.1.1 ถ้า $\sigma_1^2 = \sigma_2^2$ ค่าสถิติทดสอบที่สามารถคำนวณได้ดังนี้

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad \nu = n_1 + n_2 - 2 \quad (1)$$

โดยที่ $\bar{X}_1$ คือ ค่าเฉลี่ยตัวอย่างกลุ่มที่ 1
 s_1^2 คือ ความแปรปรวนตัวอย่างกลุ่มที่ 1
 n_1 คือ ขนาดตัวอย่างกลุ่มที่ 1
 $\bar{X}_2$ คือ ค่าเฉลี่ยตัวอย่างกลุ่มที่ 2
 s_2^2 คือ ความแปรปรวนตัวอย่างกลุ่มที่ 2
 n_2 คือ ขนาดตัวอย่างกลุ่มที่ 2
 ν คือ องศาเสรี (degree of freedom)
 s_p^2 คือ ตัวประมาณความแปรปรวนรวม โดยที่

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (2)$$

2.1.2 ถ้า $\sigma_1^2 \neq \sigma_2^2$ ค่าสถิติทดสอบที่สามารถคำนวณได้ดังนี้

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\left(\frac{1}{n_1 - 1} \right) \left(\frac{s_1^2}{n_1} \right)^2 + \left(\frac{1}{n_2 - 1} \right) \left(\frac{s_2^2}{n_2} \right)^2} \quad (3)$$

2.2 สถิติทดสอบผลบวกลำดับที่ของวิลค็อกซัน (Sinsomboonthong, 2020)

การทดสอบผลบวกลำดับที่ของวิลค็อกซันมีวิธีการคือ นำตัวอย่างทั้งสองกลุ่มมารวมเข้าด้วยกัน จากนั้นทำการเรียงลำดับตัวอย่างทั้งสองกลุ่มจากค่าน้อยที่สุดไปหาค่ามากที่สุดแล้วให้ค่าลำดับที่ โดยกำหนดให้ $y_1, y_2, \dots, y_{n_2}$ เป็นค่าที่ได้จากตัวอย่างกลุ่มที่มีขนาดน้อยกว่า (หรือเท่ากับ) ที่ถูกเรียงลำดับเรียบร้อยแล้ว และ $S_1, S_2, \dots, S_{n_2}$ คือค่าลำดับที่ของ $y_1, y_2, \dots, y_{n_2}$ ดังนั้น ค่าสถิติทดสอบผลบวกลำดับที่ของวิลค็อกซันสามารถคำนวณได้ดังนี้

$$w^* = \sum_{i=1}^{n_2} S_i \quad (4)$$

เมื่อ $n_2 \leq n_1$

3. เปรียบเทียบค่าสถิติทดสอบที่ได้ในข้อ 2 กับค่าวิกฤต (critical values) ที่ได้จากการคำนวณโดยใช้โปรแกรมทางสถิติ จากนั้นทำการสรุปผลการทดสอบตามสมมุติฐานที่กำหนดไว้ในข้อ 1 ว่าปฏิเสธหรือยอมรับสมมุติฐานว่าง (H_0)

3.1 การทดสอบที่

ถ้า $t^* \geq t_{\alpha/2, v}$ หรือ $t^* \leq -t_{\alpha/2, v}$ ปฏิเสธสมมุติฐานว่าง (H_0)

3.2 การทดสอบผลบวกลำดับที่ของวิลค็อกซัน

ถ้า $w^* \geq w_{\alpha/2}$ หรือ $w^* \leq n_2(n_1 + n_2 + 1) - w_{\alpha/2}$ ปฏิเสธสมมุติฐานว่าง (H_0)

4. ดำเนินการตามข้อ 1-3 ซ้ำ จำนวน 10,000 ครั้งในแต่ละสถานการณ์

5. นับจำนวนครั้งในการปฏิเสธสมมุติฐานว่าง

6. คำนวณค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบของทั้งสองการทดสอบ โดยค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ($\hat{\alpha}$) คำนวณได้จาก

$$\hat{\alpha} = \frac{\text{จำนวนครั้งในการปฏิเสธสมมุติฐานว่างเมื่อสมมุติฐานว่างเป็นจริง}}{\text{จำนวนครั้งในการทำซ้ำ}} \quad (5)$$

และค่าประมาณกำลังการทดสอบ ($1 - \hat{\beta}$) คำนวณได้จาก

$$1 - \hat{\beta} = \frac{\text{จำนวนครั้งในการปฏิเสธสมมุติฐานว่างเมื่อสมมุติฐานทางเลือกเป็นจริง}}{\text{จำนวนครั้งในการทำซ้ำ}} \quad (6)$$

7. เปรียบเทียบค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันกับเกณฑ์ของ Cochran (Cochran, 1947) ที่ระดับนัยสำคัญ 0.05 ถ้าค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตกอยู่ในช่วง $[0.04, 0.06]$ จะสรุปว่า การทดสอบนั้นสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้

8. เปรียบเทียบกำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซัน ถ้าการทดสอบใดสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ และให้กำลังการทดสอบที่มากกว่า จะสรุปว่า การทดสอบนั้นมีประสิทธิภาพมากกว่า

ผลการวิจัย

ผลการเปรียบเทียบประสิทธิภาพของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับประชากร 2 กลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ แสดงดังตารางที่ 1-4

ความน่าจะเป็นของความผิดพลาดแบบที่ 1

จากตารางที่ 1 และตารางที่ 2 พบว่า การทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินามและการแจกแจงปัวซองสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกสถานการณ์ แม้ในกรณีที่ตัวอย่างทั้ง 2 กลุ่มมีขนาดเล็ก ทั้งในกรณีที่ขนาดตัวอย่างทั้งสองกลุ่มเท่ากันและไม่เท่ากัน นอกจากนี้ การทดสอบทีให้ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 สูงกว่าการทดสอบผลบวกลำดับที่ของวิลค็อกซันเกือบทุกกรณี

ตารางที่ 1 ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของการทดสอบที (t-test) และการทดสอบผลบวกลำดับที่ของวิลค็อกซัน (Wilcoxon rank sum test) สำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินาม

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2)			ขนาดตัวอย่างไม่เท่ากัน (n_1, n_2)		
n	p_1	p_2		(10, 10)	(60, 60)	(100, 100)	(20, 100)	(30, 90)	(40, 80)
10	0.1	0.1	t-test	0.0500	0.0487	0.0518	0.0526	0.0560	0.0511
			Wilcoxon rank sum test	0.0477	0.0509	0.0487	0.0503	0.0526	0.0506
	0.5	0.5	t-test	0.0503	0.0511	0.0527	0.0536	0.0528	0.0527
			Wilcoxon rank sum test	0.0453	0.0505	0.0511	0.0513	0.0502	0.0512
	0.7	0.7	t-test	0.0522	0.0505	0.0529	0.0530	0.0522	0.0519
			Wilcoxon rank sum test	0.0451	0.0478	0.0520	0.0511	0.0515	0.0507

ตารางที่ 1 (ต่อ)

พารามิเตอร์			การทดสอบ	ขนาดตัวอย่าง (n_1, n_2)			ขนาดตัวอย่างไม่เท่ากัน (n_1, n_2)		
n	p_1	p_2		(10, 10)	(60, 60)	(100, 100)	(20, 100)	(30, 90)	(40, 80)
50	0.1	0.1	t-test	0.0500	0.0500	0.0539	0.0533	0.0552	0.0513
			Wilcoxon rank sum test	0.0445	0.0503	0.0552	0.0534	0.0520	0.0488
	0.5	0.5	t-test	0.0517	0.0499	0.0517	0.0526	0.0551	0.0526
			Wilcoxon rank sum test	0.0540	0.0482	0.0540	0.0530	0.0531	0.0491
	0.7	0.7	t-test	0.0499	0.0501	0.0529	0.0523	0.0541	0.0512
			Wilcoxon rank sum test	0.0427	0.0496	0.0522	0.0521	0.0518	0.0494
100	0.1	0.1	t-test	0.0498	0.0509	0.0524	0.0519	0.0553	0.0520
			Wilcoxon rank sum test	0.0429	0.0507	0.0514	0.0526	0.0529	0.0495
	0.5	0.5	t-test	0.0490	0.0485	0.0519	0.0546	0.0529	0.0459
			Wilcoxon rank sum test	0.0420	0.0509	0.0533	0.0527	0.0503	0.0503
	0.7	0.7	t-test	0.0513	0.0465	0.0532	0.0518	0.0507	0.0540
			Wilcoxon rank sum test	0.0462	0.0491	0.0521	0.0528	0.0496	0.0525

ตารางที่ 2 ค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ของการทดสอบที่ (t-test) และการทดสอบผลบวกลำดับที่ของวิลค็อกสัน (Wilcoxon rank sum test) สำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงปัวซอง

พารามิเตอร์		การทดสอบ	ขนาดตัวอย่าง (n_1, n_2)			ขนาดตัวอย่างไม่เท่ากัน (n_1, n_2)		
λ_1	λ_2		(10, 10)	(60, 60)	(100, 100)	(20, 100)	(30, 90)	(40, 80)
1	1	t-test	0.0501	0.0500	0.0524	0.0547	0.0560	0.0536
		Wilcoxon rank sum test	0.0484	0.0503	0.0509	0.0500	0.0528	0.0524
5	5	t-test	0.0486	0.0481	0.0533	0.0519	0.0544	0.0508
		Wilcoxon rank sum test	0.0431	0.0503	0.0541	0.0526	0.0523	0.0486
10	10	t-test	0.0480	0.0491	0.0499	0.0511	0.0541	0.0507
		Wilcoxon rank sum test	0.0426	0.0477	0.0494	0.0514	0.0530	0.0521

กำลังการทดสอบ

จากตารางที่ 3 และตารางที่ 4 พบว่า ในกรณีที่ขนาดตัวอย่างทั้งสองกลุ่มเท่ากัน ($n_1 = n_2$) กำลังการทดสอบของการทดสอบที่และการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินามและการแจกแจงปัวซองมีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างทั้งสองกลุ่ม และขนาดอิทธิพลเพิ่มขึ้น ดังรูปภาพ 3-4 นอกจากนี้ การทดสอบที่ให้กำลังการทดสอบสูงกว่าการทดสอบผลบวกลำดับที่ของวิลค็อกซันทุกกรณี

ตารางที่ 3 กำลังการทดสอบของการทดสอบที่ (t-test) และการทดสอบผลบวกลำดับที่ของวิลค็อกซัน (Wilcoxon rank sum test) สำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินาม

พารามิเตอร์			ขนาด อิทธิพล ^d	การทดสอบ	ขนาดตัวอย่างเท่ากัน (n_1, n_2)			ขนาดตัวอย่างไม่เท่ากัน (n_1, n_2)		
n	p_1	p_2			(10, 10)	(60, 60)	(100, 100)	(20, 100)	(30, 90)	(40, 80)
10	0.1	0.125	0.1771	t-test	0.0826	0.2720	0.4115	0.1717	0.2095	0.2476
								0.1746*	0.2215*	0.2512*
				Wilcoxon rank sum test	0.0763	0.2583	0.3939	0.1618	0.2022	0.2365
								0.1781*	0.2174*	0.2469*
	0.5	0.525	0.1119	t-test	0.0681	0.1432	0.2049	0.1018	0.1178	0.1292
								0.1007*	0.1163*	0.1299*
				Wilcoxon rank sum test	0.0612	0.1364	0.1951	0.0981	0.1115	0.1247
								0.0964*	0.1092*	0.1212*
	0.7	0.725	0.1236	t-test	0.0660	0.1658	0.2345	0.1158	0.1318	0.1537
								0.1100*	0.1288*	0.1471*
				Wilcoxon rank sum test	0.0582	0.1590	0.2285	0.1096	0.1256	0.1453
								0.1074*	0.1265*	0.1435*
50	0.1	0.125	0.3959	t-test	0.2190	0.8526	0.9729	0.6146	0.7443	0.8140
								0.6190*	0.7517*	0.8136*
				Wilcoxon rank sum test	0.1978	0.8358	0.9667	0.5964	0.7249	0.7923
								0.5893*	0.7265*	0.7948*
	0.5	0.525	0.2502	t-test	0.1219	0.4816	0.7065	0.2966	0.3822	0.4395
								0.2941*	0.3824*	0.4405*
				Wilcoxon rank sum test	0.1033	0.4652	0.6892	0.2826	0.3616	0.4174
								0.2831*	0.3645*	0.4192*

ตารางที่ 3 (ต่อ)

พารามิเตอร์			ขนาด อิทธิพล ^d	การทดสอบ	ขนาดตัวอย่างเท่ากัน (n_1, n_2)			ขนาดตัวอย่างไม่เท่ากัน (n_1, n_2)		
n	p_1	p_2			(10, 10)	(60, 60)	(100, 100)	(20, 100)	(30, 90)	(40, 80)
100	0.7	0.725	0.2763	t-test	0.1311	0.5698	0.7925	0.3570	0.4549	0.5137
								0.3516*	0.4482*	0.5198*
				Wilcoxon rank	0.1143	0.5508	0.7777	0.3359	0.4334	0.4937
	0.5	0.525	0.3538	sum test				0.3401*	0.4290*	0.4947*
				t-test	0.3879	0.9894	0.9998	0.8870	0.9596	0.9818
								0.8871*	0.9570*	0.9824*
100	0.7	0.725	0.3907	Wilcoxon rank	0.3517	0.9873	0.9995	0.8756	0.9507	0.9777
				sum test				0.8678*	0.9494*	0.9770*
				t-test	0.1854	0.7740	0.9449	0.5317	0.6536	0.7243
	0.5	0.525	0.3538					0.5266*	0.6529*	0.7218*
				Wilcoxon rank	0.1619	0.7507	0.9329	0.5078	0.6273	0.7044
				sum test				0.5039*	0.6280*	0.6983*
100	0.7	0.725	0.3907	t-test	0.2065	0.8491	0.9741	0.6052	0.7385	0.8034
								0.5994*	0.7299*	0.8050*
				Wilcoxon rank	0.1893	0.8303	0.9671	0.5796	0.7140	0.7840
	0.5	0.525	0.3538	sum test				0.5797*	0.7054*	0.7793*
				t-test	0.2065	0.8491	0.9741	0.6052	0.7385	0.8034
								0.5994*	0.7299*	0.8050*

* ค่าประมาณกำลังการทดสอบของขนาดตัวอย่าง (n_1, n_2) เท่ากับ (100, 20) (90, 30) และ (80, 40)

^d ขนาดอิทธิพล (Effect size) คำนวณจาก $\left| (np_1 - np_2) / \sqrt{np_1(1-p_1) + np_2(1-p_2)} \right|$

เมื่อพิจารณาอิทธิพลของอัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ที่มีต่อกำลังการทดสอบภายใต้ขนาดตัวอย่างรวมที่เท่ากัน พบว่า สำหรับข้อมูลที่มีการแจกแจงทวินาม โดยที่ $p_1 < p_2$ เมื่ออัตราส่วนของ $n_1 : n_2$ เท่ากับ 1 : 1 ยกเว้นกรณีที่ n_1 และ n_2 เท่ากับ 10 กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันจะมีค่าสูงสุด และเมื่ออัตราส่วนของ $n_1 : n_2$ ลดลงเป็น 1 : 2, 1 : 3 และ 1 : 5 กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันจะลดลงตามลำดับ ดังรูปภาพ 5-6 ในทำนองเดียวกัน ถ้า $p_1 < p_2$ และ $n_1 > n_2$ พบว่า เมื่ออัตราส่วนของ $n_1 : n_2$ เพิ่มขึ้นจาก 1 : 1 เป็น 2 : 1, 3 : 1 และ 5 : 1 กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันก็จะลดลงตามลำดับ

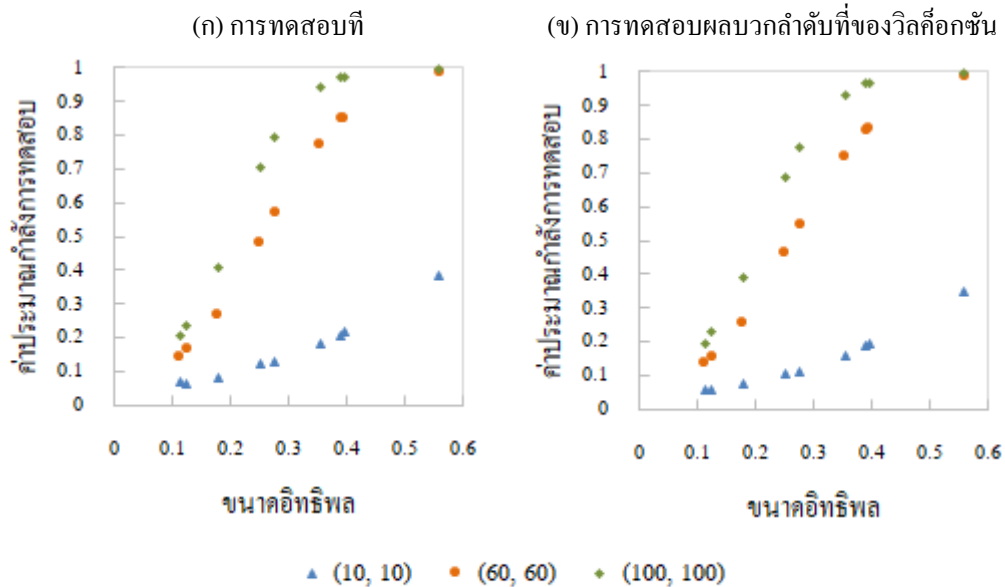
ตารางที่ 4 กำลังการทดสอบของการทดสอบที (t-test) และการทดสอบผลบวกลำดับที่ของวิลค็อกซัน (Wilcoxon rank sum test) สำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงปัวซง

พารามิเตอร์ λ_1	ขนาด λ_2 อิทธิพล ^d		การทดสอบ	ขนาดตัวอย่างเท่ากัน (n_1, n_2)			ขนาดตัวอย่างไม่เท่ากัน (n_1, n_2)		
				(10, 10)	(60, 60)	(100, 100)	(20, 100)	(30, 90)	(40, 80)
1	1.25	0.1667	t-test	0.0766	0.2418	0.3728	0.1624	0.1970	0.2284
							0.1538*	0.1914*	0.2270*
			Wilcoxon rank sum test	0.0740	0.2287	0.3488	0.1453	0.1803	0.2090
1	1.50	0.3162	t-test	0.1567	0.6765	0.8772	0.4283	0.5505	0.6329
							0.4171*	0.5448*	0.6206*
			Wilcoxon rank sum test	0.1454	0.6446	0.8500	0.4062	0.5153	0.5999
1	2	0.5774	t-test	0.4048	0.9937	0.9999	0.9104	0.9737	0.9893
							0.8817*	0.9630*	0.9861*
			Wilcoxon rank sum test	0.3780	0.9910	0.9998	0.9000	0.9656	0.9848
5	8	0.8321	t-test	0.7068	1.000	1.000	0.9982	0.9999	1.000
							0.8679*	0.9999*	1.000*
			Wilcoxon rank sum test	0.6707	1.000	1.000	0.9974	0.9997	1.000
5	8	0.8321	t-test	0.7068	1.000	1.000	0.9982	0.9999	1.000
							0.8679*	0.9999*	1.000*
			Wilcoxon rank sum test	0.6707	1.000	1.000	0.9974	0.9997	1.000

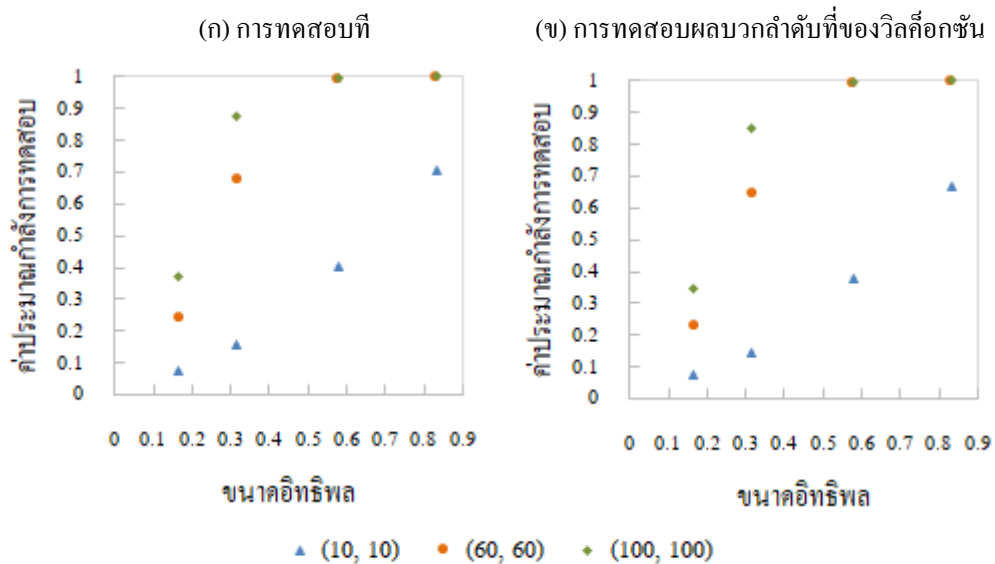
* ค่าประมาณกำลังการทดสอบของขนาดตัวอย่าง (n_1, n_2) เท่ากับ (100, 20) (90, 30) และ (80, 40)

^d ขนาดอิทธิพล (Effect size) คำนวณจาก $|(\lambda_1 - \lambda_2) / \sqrt{\lambda_1 + \lambda_2}|$

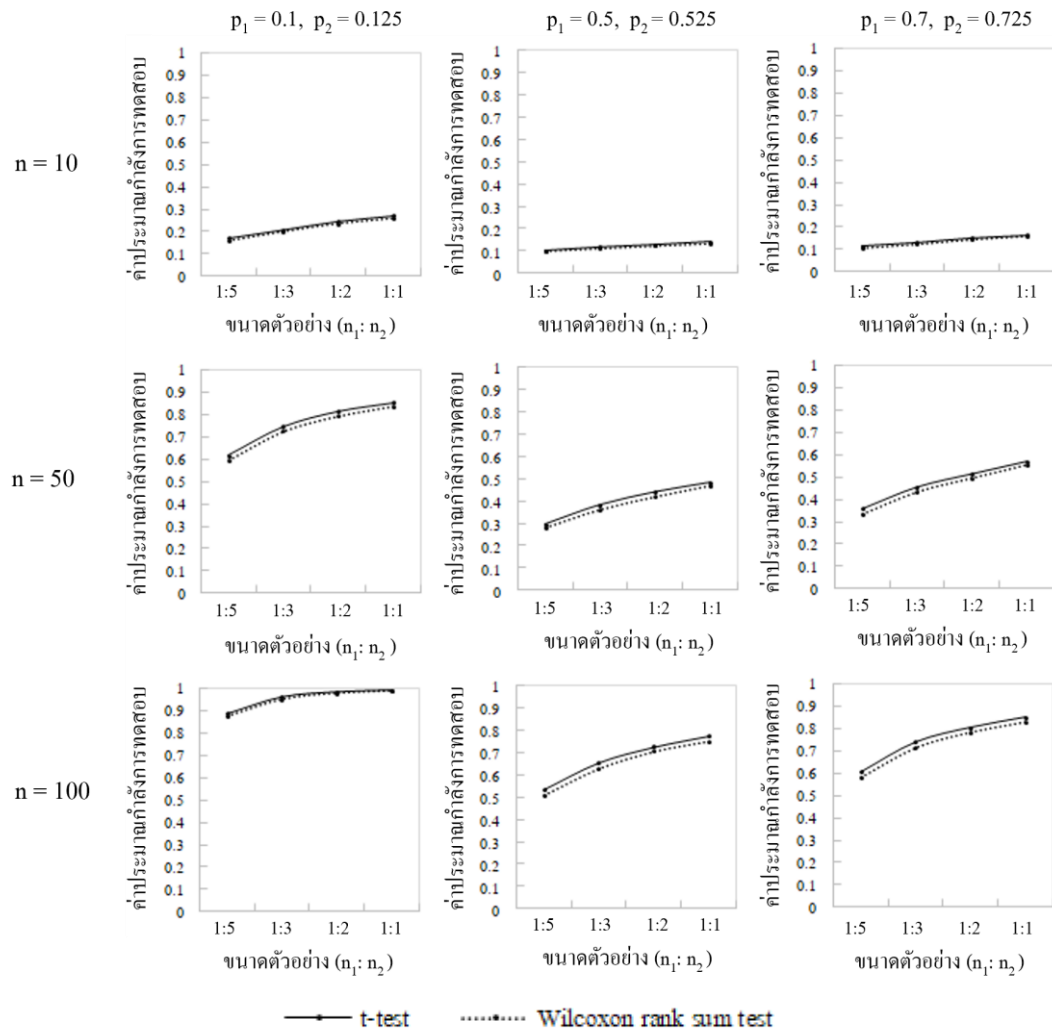
สำหรับข้อมูลที่มีการแจกแจงปัวซง เมื่อ $\lambda_1 < \lambda_2$ และ $n_1 < n_2$ (หรือ $n_1 > n_2$) พบว่า กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันจะมีค่าสูงสุดเมื่ออัตราส่วนของ $n_1 : n_2$ เท่ากับ 1 : 1 เช่นเดียวกับข้อมูลที่มีการแจกแจงทวินาม นอกจากนี้ การทดสอบทีให้กำลังการทดสอบสูงกว่าการทดสอบผลบวกลำดับที่ของวิลค็อกซันทุกกรณี



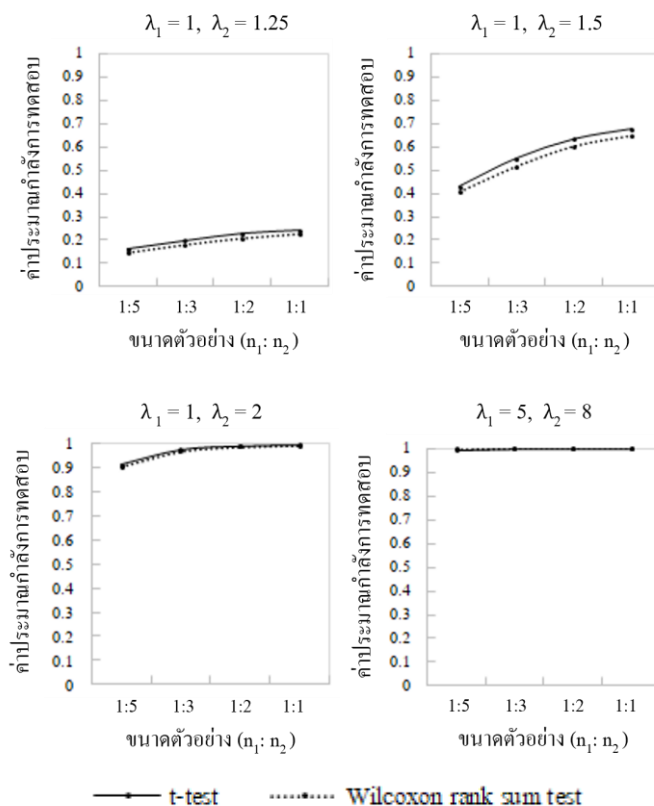
รูปที่ 3 ความสัมพันธ์ระหว่างขนาดอิทธิพลและค่าประมาณกำลังการทดสอบสำหรับข้อมูลที่มีการแจกแจงทวินาม



รูปที่ 4 ความสัมพันธ์ระหว่างขนาดอิทธิพลและค่าประมาณกำลังการทดสอบสำหรับข้อมูลที่มีการแจกแจงปัวซอง



รูปที่ 5 ค่าประมาณกำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซัน สำหรับทดสอบค่ากลางของข้อมูลที่มีการแจกแจงทวินาม กรณีขนาดตัวอย่าง 2 กลุ่มไม่เท่ากัน



รูปที่ 6 ค่าประมาณกำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซัน สำหรับการทดสอบค่ากลางของข้อมูลที่มีการแจกแจงปัวซง กรณีขนาดตัวอย่าง 2 กลุ่มไม่เท่ากัน

สรุปและวิจารณ์ผลการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของประชากรสองกลุ่มที่เป็นอิสระกันเมื่อข้อมูลเป็นจำนวนนับ โดยการแจกแจงที่นำมาศึกษา คือ การแจกแจงทวินาม และการแจกแจงปัวซง การเปรียบเทียบประสิทธิภาพของการทดสอบพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ โดยกำหนดระดับนัยสำคัญที่ 0.05

สำหรับความสามารถในการควบคุมความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินามและการแจกแจงปัวซงพบว่า การทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทุกกรณีที่ศึกษา แม้ในกรณีที่ตัวอย่างทั้ง 2 กลุ่มมีขนาดเล็ก

และในกรณีที่ขนาดตัวอย่างทั้งสองกลุ่มเท่ากันและไม่เท่ากัน ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของ Araveeporn et al. (2017) ที่ได้ทำการศึกษาประสิทธิภาพของการทดสอบซี (z-test) การทดสอบที (t-test) การทดสอบแบบสุ่ม (randomization test) และการทดสอบแมนน์-วิทนียู (Mann-Whitney U test) สำหรับทดสอบค่าเฉลี่ยหรือค่ากลางของประชากร 2 กลุ่มที่เป็นอิสระกัน โดยผลการวิจัยพบว่า การทดสอบซี การทดสอบที และการทดสอบแมนน์-วิทนียูสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ทั้งในกรณีตัวอย่างขนาดเล็กและตัวอย่างขนาดใหญ่

เมื่อพิจารณากำลังการทดสอบพบว่า ในกรณีที่ขนาดตัวอย่างทั้งสองกลุ่มเท่ากัน กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันสำหรับทดสอบความแตกต่างระหว่างค่ากลางของข้อมูลที่มีการแจกแจงทวินามและการแจกแจงปัวซองมีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างทั้งสองกลุ่ม และขนาดอิทธิพลเพิ่มขึ้น เมื่อพิจารณาอิทธิพลของอัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ที่มีต่อกำลังการทดสอบภายใต้ขนาดตัวอย่างรวมที่เท่ากันพบว่า กำลังการทดสอบของการทดสอบทีและการทดสอบผลบวกลำดับที่ของวิลค็อกซันจะมีค่าสูงสุดเมื่ออัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ($n_1 : n_2$) เท่ากับ 1 : 1 ยกเว้นกรณีที่ n_1 และ n_2 เท่ากับ 10 และจะมีค่าน้อยที่สุดเมื่ออัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ($n_1 : n_2$) เท่ากับ 1 : 5 (หรือ 5 : 1) ทั้งในกรณีที่ข้อมูลมีการแจกแจงทวินามและกรณีที่ข้อมูลมีการแจกแจงปัวซอง ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของ Kim & Park (2019) ที่ได้ทำการศึกษาเกี่ยวกับข้อตกลงเบื้องต้นของการทดสอบทีสำหรับประชากร 2 กลุ่มที่เป็นอิสระกันและพบว่า ถ้าอัตราส่วนของขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ($n_1 : n_2$) เท่ากับ 1 : 1 จะให้กำลังการทดสอบที่สูงที่สุด คือ 0.88 รองลงมาคืออัตราส่วนของ $n_1 : n_2$ เท่ากับ 5 : 3, 3 : 1, และ 7 : 1 ซึ่งให้กำลังการทดสอบ 0.86, 0.78, และ 0.55 ตามลำดับ นอกจากนี้ในการวิจัยครั้งนี้ยังพบว่า การทดสอบทีซึ่งเป็นการทดสอบแบบอิงพารามิเตอร์ให้กำลังการทดสอบสูงกว่าการทดสอบผลบวกลำดับที่ของวิลค็อกซันทุกกรณี

จากผลการวิจัยในครั้งนี้จะเห็นได้ว่า สำหรับข้อมูลที่เป็นจำนวนนับที่มีการแจกแจงทวินามและการแจกแจงปัวซอง การทดสอบทีเป็นการทดสอบความแตกต่างระหว่างค่ากลางของประชากรสองกลุ่มที่เป็นอิสระกันที่มีประสิทธิภาพดีกว่าการทดสอบวิลค็อกซัน เนื่องจากสามารถควบคุมความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ได้ และมีกำลังการทดสอบที่สูงกว่าถึงแม้ว่าข้อมูลจะไม่เป็นไปตามข้อตกลงเบื้องต้นของการทดสอบที กล่าวคือข้อมูลไม่ได้มาจากประชากรที่มีการแจกแจงปกติและไม่ได้มีมาตรการควบคุมตัวอยู่ในมาตราช่วงเป็นอย่างน้อย ทั้งนี้สาเหตุอาจเนื่องมาจากการทดสอบทีเป็นการทดสอบที่มีความแข็งแกร่งต่อการละเมิดข้อตกลงเบื้องต้น

เอกสารอ้างอิง

Araveeporn, A., Chuenarrom, C., Fuksab, Supawee Choedchitkusol, S., & Muangmeesup, A. (2017). An efficiency comparison of z-test, t-test, the randomization test and Mann-Whitney U test for testing two independent population means or medians. *Journal of Science Ladkrabang*, 26(2), 73-88.

- Blair, R. C., & Higgins, J. J. (1980). A comparison of the power of the Wilcoxon's rank-sum statistic to that of Student's t statistic under various non-normal distributions. *Journal of Educational Statistics*, 5(4), 309–335.
- Bridge, P. D., & Sawilowsky, S. S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and Wilcoxon rank sum rank-sum test in small samples applied research. *Journal of clinical epidemiology*, 52(3), 229-235.
- Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 3, 22-38.
- Kim, T. K., & Park, J. H. (2019). More about the basic assumptions of t-test: normality and sample size. *Korean journal of anesthesiology*, 72(4), 331-335.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2014). *Introduction to the Practice of Statistics* (8th ed.). New York, USA: W. H. Freeman and Company.
- Sinsomboonthong, S. (2020). *Nonparametric Statistics*. Bangkok, Thailand: Chamchuree Products Co., LTD.
- Sangthong, M. (2013). A study of efficiency of parametric and nonparametric statistics in testing of central difference between two populations. *KKU Science Journal*, 41(1), 226-238.
- Sangthong, M. (2014). Robustness and power of the test of parametric and nonparametric statistics in testing of central difference between two populations for Likert-type data 5 point. *Thai Science and Technology Journal*, 22(5), 605-619.