ว.วิทย. มข. 41(1) 226-238 (2556) KKU Sci. J. 41(1) 226-238 (2013)

การศึกษาประสิทธิภาพของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่ม

A Study of Efficiency of Parametric and Nonparametric Statistics in Testing of Central Difference between Two Populations

มนตรี สังข์ทอง[1]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่ม ซึ่งตัวสถิติทดสอบที่นำมาศึกษาประกอบด้วย t test, Mann Whitney U test, Van der Waerden test, Wald Wolfowitz Runs test และ Modified U test โดยจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล จำแนกสถานการณ์โดยกำหนดประชากรทั้งที่มีการแจกแจงปกติ การแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ และการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ กำหนดขนาดตัวอย่างจากประชากร 2 กลุ่ม เท่ากับ (5,5), (5,10) เป็นตัวแทนกลุ่มตัวอย่างขนาดเล็ก (20,20), (25,20) เป็นตัวแทนกลุ่มตัวอย่างขนาดกลาง และ (50,50), (50,100) เป็นตัวแทนกลุ่มตัวอย่างขนาดใหญ่ กำหนดอัตราส่วนของความแปรปรวน เท่ากับ (1:1) และ (1:2) ณ ระดับนัยสำคัญ 0.05 โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ คือ ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบ ผลการวิจัย พบว่า เมื่อประชากรมีการแจกแจงปกติ ตัวสถิติทดสอบที่มีกำลังของการทดสอบสูงสุดและสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ t test และ Van der Waerden test เมื่อกลุ่มตัวอย่างมีขนาดกลางและขนาดใหญ่ คือ Modified U test เมื่อประชากรมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ พบว่า ตัวสถิติทดสอบที่มีอำนาจการทดสอบสูงสุดและสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ t test และ Van der Waerden test เมื่อกลุ่มตัวอย่างมีขนาดกลางและขนาดใหญ่ คือ Mann Whitney U test และเมื่อประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ พบว่า ตัวสถิติทดสอบที่มีอำนาจการทดสอบสูงสุดและสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ t test และ Van der Waerden test เมื่อกลุ่มตัวอย่างมีขนาดกลาง คือ Mann Whitney U test และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ t test และ Mann Whitney U test

[1]สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.พระนครศรีอยุธยา 13000

E-mail: sonktong@yahoo.com

ABSTRACT

This research aimed to study the efficiency of parametric and nonparametric statistics in testing of central difference between two populations. The test statistics used to study the efficiency were t test, Mann Whitney U test, Van der Waerden test, Wald Wolfowitz Runs test and Modified U test. Classification of the population according to normal distribution, negative skewness and leptokurtic kurtosis distribution and positive skewness and leptokurtic kurtosis distribution. The sample sizes of two population were (5,5), (5,10), a small sample representative, (20,20), (25,20), a medium sample representative, and (50,50), (50,100), a large sample representative. The ratios of variance were (1:1) and (1:2) at a significant level of 0.05. The criteria used to compare the efficiency were the ability to control the type I error and power of the test. The results showed that: When the population has normal distribution, The test statistics have the highest power of the test and to control the type I error for small sample sizes were t test and Van der Waerden test, for medium and large sample sizes were Modified U test. When the population has negative skewness and leptokurtic kurtosis distribution, The test statistics have the highest power of the test and to control the type I error for small sample sizes were t test and Van der Waerden test, for medium and large sample sizes were Mann Whitney U test. When the population has positive skewness and leptokurtic kurtosis distribution, The test statistics have the highest power of the test and to control the type I error for small sample sizes were t test and Van der Waerden test, for medium sample sizes were Mann Whitney U test and for large sample sizes were t test and Mann Whitney U test.

คำสำคัญ: การทดสอบความแตกต่างของค่ากลาง ความคลาดเคลื่อนประเภทที่ 1 กำลังของการทดสอบ
Keywords: The central difference test, Type I Error, Power of the test

บทนำ

การวิจัยเป็นกระบวนการแสวงหา ศึกษา ค้นคว้า หรือพัฒนาองค์ความรู้ใหม่ ๆ ที่มีลักษณะเป็นนัยทั่วไปอย่างมีระบบแบบแผน โดยใช้กระบวนการทางวิทยาศาสตร์ ซึ่งประกอบไปด้วยขั้นตอนที่สำคัญ ได้แก่ การกำหนดประเด็นปัญหา การตั้งสมมติฐาน การเก็บรวบรวมข้อมูล การวิเคราะห์ และการทดสอบสมมติฐาน สรุปผลและยืนยันข้อค้นพบ ดังนั้นสถิติจึงเป็นเทคนิคที่เกี่ยวข้องกับการวิจัยโดยตรง นักวิจัยต้องรู้จักนำสถิติมาประยุกต์ใช้เป็นเครื่องมือของการวิจัย ซึ่งหมายความว่า นักวิจัยต้องมีความรู้ความสามารถในการวิเคราะห์ข้อมูล มีการสรุปอ้างอิงที่เชื่อถือได้ (ศิริชัย, 2545) จากคำกล่าวข้างต้นแสดงให้เห็นว่าสถิติเป็นวิธีการ (method) ซึ่งเป็นส่วนหนึ่งของระเบียบวิธีวิจัย (research methodology) ที่สำคัญในการสรุปอ้างอิงไปถึงค่าพารามิเตอร์ (parameter) ของประชากรโดยอาศัยข้อมูลจากการชักตัวอย่าง ดั้งนั้นกระบวนการในการเลือกสถิติทดสอบที่เหมาะสมกับแต่ละ

สถานการณ์ของข้อมูลจึงเป็นสิ่งที่สำคัญอย่างยิ่ง โดยต้องคำนึงถึงข้อตกลงเบื้องต้น (basic assumption) ของตัวสถิติทดสอบแต่ละชนิดด้วย

ในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างของค่ากลางระหว่างประชากรสองกลุ่มที่เป็นอิสระกัน ตัวสถิติทดสอบที่ได้รับความนิยมและนำไปใช้อย่างแพร่หลาย คือ two sample student t test โดยมีข้อตกลงเบื้องต้นว่าประชากรต้องมีการแจกแจงปกติ และมีความแปรปรวนเท่ากัน แม้ว่า two sample student t test จะเป็นตัวสถิติทดสอบที่มีความแกร่งเมื่อข้อตกลงเบื้องต้นถูกละเลย (Boneau, 1960) ทั้งนี้ ถ้ามีการละเลยข้อตกลงเกี่ยวกับความแปรปรวนในกรณีที่กลุ่มตัวอย่างไม่เท่ากันจะมีผลกระทบอย่างมาก(Hsu, 1938; Scheffe, 1959, 1970; Zimmerman and Zumbo, 2009) จากปัญหาดังกล่าวนักสถิติจึงได้พัฒนาตัวสถิติทดสอบที่เรียกว่า Welch t test ซึ่งเป็นสถิติอิงพารามิเตอร์แต่ไม่มีข้อตกลงเกี่ยวกับความแปรปรวนของประชากร จากการศึกษาของ Zimmerman and Zumbo (2009) พบว่า Welch t test มีค่าความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับค่าระดับนัยสำคัญทางสถิติที่กำหนดมากที่สุด เมื่อประชากรมีการแจกแจงปกติ และขนาดตัวอย่างแตกต่างกันมาก

การทดสอบสมมติฐานดังกล่าวถ้าข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น ผู้วิจัยสามารถทดสอบโดยเลือกใช้สถิติไม่อิงพารามิเตอร์ ซึ่งมีหลายตัวสถิติทดสอบ เช่น Mann-Whitney U test เป็นต้น ซึ่งตัวสถิติทดสอบดังกล่าวสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ และมักจะมีกำลังของการทดสอบสูงกว่า t test เมื่อประชากรไม่มีการแจกแจงปกติ จากการศึกษาของนักสถิติที่ผ่านมา พบว่า การเปลี่ยนแปลงของระดับนัยสำคัญในการทดสอบไม่อิงพารามิเตอร์เป็นผลกระทบจากการที่มีความแปรปรวนไม่เท่ากัน พร้อม ๆ กับที่มีขนาดตัวอย่างไม่เท่ากัน อย่างไรก็ตามพบว่าคุณสมบัติของการทดสอบด้วยสถิติไม่อิงพารามิเตอร์บางตัวสถิติทดสอบเมื่ออยู่ในสถานการณ์ต่าง ๆ ยังไม่ชัดเจน

จากที่กล่าวมาข้างต้นทำให้ผู้วิจัยสนใจที่จะศึกษาประสิทธิภาพของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่ม โดยตัวสถิติทดสอบที่ศึกษาเปรียบเทียบกัน ได้แก่ t test, Mann Whitney U test, Wald Wolfowitz Runs test, Van der Waerden test และ Modified U test โดยจำลองข้อมูลในสถานการณ์ต่าง ๆ เพื่อให้ได้สารสนเทศประกอบการตัดสินใจในการเลือกใช้ตัวสถิติทดสอบที่เหมาะสมในแต่ละสถานการณ์ และได้สารสนเทศพื้นฐานในการพัฒนาตัวสถิติทดสอบที่มีความแกร่ง (robust) ต่อไป

วิธีการทางสถิติ

การศึกษาประสิทธิภาพของสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่ม สำหรับตัวสถิติทดสอบต่าง ๆ ที่นำมาศึกษาเปรียบเทียบกัน มีรายละเอียดดังนี้

2.1 t test

t test เป็นสถิติอิงพารามิเตอร์ที่ได้รับความนิยมในการนำไปใช้อย่างแพร่หลาย โดยแบ่งออกเป็น 2 กรณี คือ two sample student t test เมื่อ $\sigma_1^2 = \sigma_2^2$ และ Welch t test เมื่อ $\sigma_1^2 \neq \sigma_2^2$ มีสูตรในการคำนวณดังนี้ (Hopkins et al., 1987)

กรณี $\sigma_1^2 = \sigma_2^2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

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

$$df = n_1 + n_2 - 2$$

กรณี $\sigma_1^2 \neq \sigma_2^2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$df = \frac{\left[\left(s_1^2 / n_1\right) + \left(S_2^2 / n_2\right)\right]^2}{\frac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}}$$

จะปฏิเสธ H_0 เมื่อ $t > t_{\frac{\alpha}{2}}$ หรือ $t < -t_{\frac{\alpha}{2}}$

สำหรับในการวิจัยครั้งนี้ดำเนินการทดสอบความเท่ากันของความแปรปรวนด้วย Levene’s test ถ้าความแปรปรวนเท่ากันทดสอบผลต่างของค่าเฉลี่ยด้วย two sample student t test เมื่อความแปรปรวนไม่เท่ากันทดสอบผลต่างของค่าเฉลี่ยด้วย Welch t test

2.2 Mann Whitney U test

Mann และ Whitney ได้คิดค้นวิธีการทดสอบนี้เมื่อปี ค.ศ. 1947 และให้ค่าสถิติที่คำนวณได้ เป็นค่า U จึงเรียกการทดสอบนี้ว่า “Mann Whitney U test” ซึ่งใช้อันดับ (rank) ของข้อมูลมาช่วยในการคำนวณมีสูตรในการคำนวณ ดังนี้ (Sheskin, 2000)

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \sum R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum R_2$$

$$U = Min(U_1, U_2)$$

เมื่อกลุ่มตัวอย่างมีขนาดใหญ่จะมีลักษณะใกล้เคียงการแจกแจงปกติ จะนำ U Statistics มาประมาณด้วย z โดยมีสูตรการคำนวณ ดังนี้

$$z = \frac{U - \mu_U}{\sigma_u}$$

เมื่อ $\mu_U = \dfrac{n_1 n_2}{2}$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

จะปฏิเสธ H_0 เมื่อ $U < u_0$ โดยที่ $P(U < u_0) = \dfrac{\alpha}{2}$ หรือเมื่อกลุ่มตัวอย่างมีขนาดใหญ่สามารถกำหนดค่าวิกฤต แทนด้วย $U > z_{\frac{\alpha}{2}}$ หรือ $U < -z_{\frac{\alpha}{2}}$

2.3 Van der Waerden test

Van der Waerden test เป็นตัวสถิติทดสอบที่ถูกค้นพบโดย นักคณิตศาสตร์ชาวดัตช์ ชื่อว่า Bartel Leendert van der Waerden ในปี 1952 ซึ่งสามารถใช้ทดสอบความแตกต่างระหว่างค่ากลาง 2 กลุ่ม ซึ่งมีรายละเอียดการคำนวณ ดังนี้ (Van der Waerden, B.L. 1952)

$$T_1 = \frac{1}{s^2}\sum_{i=1}^{2} n_1 \overline{A}_i^2$$

เมื่อ $A_{ij} = \Phi^{-1}\left(\dfrac{R(X_{ij})}{N+1}\right)$ คือ คะแนนมาตรฐานตัวที่ j กลุ่มที่ i

$\overline{A}_i = \dfrac{1}{n_i}\displaystyle\sum_{j=1}^{n_1} A_{ij}$ i = 1,2 คือ ค่าเฉลี่ยของคะแนนมาตรฐานกลุ่มที่ i

$s^2 = \dfrac{1}{N-1}\displaystyle\sum_{i=1}^{2}\sum_{j=1}^{n_i} A_{ij}^2$ คือ ค่าความแปรปรวนของคะแนนมาตรฐาน

จะปฏิเสธ H_0 เมื่อ $T_1 > \chi^2_{\alpha,1}$

2.4 Wald Wolfowitz Runs test

Abraham Wald and Jacob Wolfowitz (1940) ได้นำเสนอตัวสถิติทดสอบที่เรียกว่า Wald Wolfowitz Runs test ซึ่งสามารถใช้ทดสอบความแตกต่างระหว่างค่ากลาง 2 กลุ่ม ซึ่งมีรายละเอียดการคำนวณ ดังนี้

กรณีที่จำนวนค่าสังเกตไม่เท่ากัน

r คือ จำนวนรันทั้งหมดหรือจำนวนครั้งของการเปลี่ยนอันดับของแต่ละกลุ่มเมื่อเรียงอันดับจากน้อยไปหามากหรือมากไปหาน้อย

กรณีที่จำนวนค่าสังเกตเท่ากัน จะคำนวณค่า r ได้ดังนี้

$$r = \frac{r_1 + r_2}{2}$$

เมื่อ r_1 คือค่าต่ำสุดของจำนวนรันของข้อมูลที่เรียงอันดับร่วมกัน 2 กลุ่ม โดยนำค่าที่เท่ากันของข้อมูลกลุ่มเดียวกับข้อมูลค่าก่อนหน้ามาเรียงอันดับ

r_2 คือ ค่าต่ำสุดของจำนวนรันของข้อมูลที่รียงอันดับร่วมกัน 2 กลุ่ม โดยนำค่าที่เท่ากันของข้อมูลคนละกลุ่มเดียวกับข้อมูลค่าก่อนหน้ามาเรียงอันดับ

ถ้าตัวอย่างมีขนาดใหญ่ จากทฤษฎีลิมิตเข้าสู่ส่วนกลาง (central limit theorem) การแจกแจงของ r จะมีการแจกแจงโดยประมาณใกล้เคียงกับการแจกแจงปกติ จะประมาณด้วย z สามารถคำนวณได้ดังนี้

$$z = \frac{r - \mu_r}{\sigma_r}$$

เมื่อ $$\mu_r = \frac{2n_1 n_2}{n} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n-1)}}$$

$$n = n_1 + n_2$$

จะปฏิเสธ H_0 เมื่อ $z > z_{\frac{\alpha}{2}}$ หรือ $z < -z_{\frac{\alpha}{2}}$

2.5 Modified U test

Fligner and Policello (1981) ได้เสนอการทดสอบความแตกต่างระหว่างประชากรสองกลุ่ม โดยไม่ต้องมีข้อสมมติว่าประชากรมีความแปรปรวนเท่ากันเพื่อแก้ปัญหาที่เรียกว่า ปัญหาของบีเรนส์-ฟิชเชอร์ (Behrens-Fisher problem) โดยมีรายละเอียดดังนี้

กรณีที่จำนวนค่าสังเกตไม่เท่ากัน

กำหนดให้ P_i คือ จำนวนค่าสังเกตในกลุ่มตัวอย่างที่ 2$\left(Y_j\right)$ ที่น้อยกว่าจำนวนค่าสังเกตในกลุ่มตัวอย่างที่ 1 $\left(X_i\right)$

Q_j คือ จำนวนค่าสังเกตในกลุ่มตัวอย่างที่ 1$\left(X_i\right)$ ที่น้อยกว่าจำนวนค่าสังเกตในกลุ่มตัวอย่างที่ 2 $\left(Y_j\right)$

กรณีที่จำนวนค่าสังเกตเท่ากัน

กำหนดให้ P_i คือ (จำนวน Y_j ที่น้อยกว่า X_i) + $\frac{1}{2}$ (จำนวน Y_j ที่น้อยกว่า X_i)

Q_j คือ (จำนวน X_i ที่น้อยกว่า Y_j) + $\frac{1}{2}$ (จำนวน X_i ที่น้อยกว่า Y_j)

สามารถคำนวณค่าสถิติทดสอบได้ดังนี้

$$\hat{U} = \frac{\sum_{j=1}^{n_2} Q_j - \sum_{i=1}^{n_1} P_i}{2\sqrt{V_1 + V_2 + \overline{P}\,\overline{Q}}}$$

เมื่อ $$\overline{P} = \frac{\sum_{i=1}^{n_1} P_i}{n_1}$$

$$\overline{Q} = \frac{\sum_{j=1}^{n_2} Q_j}{n_2}$$

$$V_1 = \sum_{i=1}^{n_1} \left(P_i - \overline{P}\right)^2$$

$$V_2 = \sum_{j=1}^{n_2} \left(Q_j - \overline{Q}\right)^2$$

จะปฏิเสธ H_0 เมื่อ $\hat{U} > u_{\frac{\alpha}{2}}$ หรือ $\hat{U} < -u_{\frac{\alpha}{2}}$ หรือเมื่อกลุ่มตัวอย่างมีขนาดใหญ่สามารถกำหนดค่าวิกฤต แทนด้วย $\hat{U} > z_{\frac{\alpha}{2}}$ หรือ $\hat{U} < -z_{\frac{\alpha}{2}}$

วิธีการวิจัย

การวิจัยครั้งนี้ดำเนินการจำลองแบบปัญหาโดยใช้เทคนิคมอนติคาร์โลด้วยโปรแกรมภาษา R มีขั้นตอนดังนี้

1. ประชากรที่ศึกษาประกอบด้วยประชากรที่มีการแจกแจงปกติ ประชากรที่มีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ (S_k = -0.65, K = 6.00) และประชากรที่มีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ (S_k = 0.65, K = 6.00)

การสร้างประชากรที่มีการแจกแจงดังกล่าว จะใช้วิธีของ Ramberg และคณะ ซึ่งเสนอวิธีการสร้างตัวแปรสุ่มที่ขึ้นอยู่กับความเบ้ (skewness) และความโด่ง (kurtosis) ซึ่งตัวแปรสุ่มนี้ถูกกำหนดจากค่าพารามิเตอร์ 4 ค่า ดังนี้

$$x = \lambda_1 + \frac{[R^{\lambda_3} - (1-R)^{\lambda_4}]}{\lambda_2}$$

โดยที่ R เป็นตัวเลขสุ่มที่มีการแจกแจงเอกรูปในช่วง (0,1)

λ_1 เป็นพารามิเตอร์ตำแหน่ง (location parameter)

λ_2 เป็นพารามิเตอร์มาตรา (scale parameter)

λ_3, λ_4 เป็นพารามิเตอร์รูปร่าง (shape parameter) ซึ่งขึ้นกับค่าความเบ้และความโด่งที่กำหนด ถ้าการแจกแจงสมมาตร จะได้ $\lambda_3 = \lambda_4$

สำหรับค่า $\lambda_1, \lambda_2, \lambda_3$ และ λ_4 อ่านจากตารางการแจกแจงของ Ramberg et al. (1979) ตามค่าความเบ้และค่าความโด่งที่กำหนด โดยค่า λ_1, λ_2 ที่ได้มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 แต่ถ้าค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 จะต้องแปลงค่า λ_1, λ_2 จากตาราง ดังนี้

$$\lambda_1(\mu, \sigma^2) = \lambda_1(0,1)\ \sigma + \mu$$

$$\lambda_2(\mu, \sigma^2) = \lambda_2(0,1) / \sigma$$

ในการวิจัยครั้งนี้กำหนดค่าพารามิเตอร์ของประชากรกลุ่มที่ 1 คือ ค่าเฉลี่ย เท่ากับ 10 และค่าความแปรปรวน เท่ากับ 3 สำหรับประชากรกลุ่มที่ 2 แปรเปลี่ยนตามเงื่อนไขที่กำหนด โดยในการศึกษาครั้งนี้ทดสอบสมมติฐานแบบสองหาง

2. ขนาดตัวอย่างที่ทำการศึกษาแบ่งออกเป็น 6 ขนาด ได้แก่ 5 10 20 25 50 และ 100 โดยกำหนดให้ขนาดตัวอย่างเท่ากับ 5 และ 10 เป็นตัวแทนกลุ่มตัวอย่างขนาดเล็ก ขนาดตัวอย่าง 20 และ 25 เป็นตัวแทนกลุ่มตัวอย่างขนาดกลาง และขนาดตัวอย่าง 50 และ 100 เป็นตัวแทนของกลุ่มตัวอย่างขนาดใหญ่ โดยมีรายละเอียดดังนี้

n_1	5	5	20	25	50	50
n_2	5	10	20	20	50	100

3. กำหนดอัตราส่วนค่าเฉลี่ยของประชากร จำนวน 2 ขนาด คือ 1:1 และ 1:1.2

4. กำหนดอัตราส่วนค่าความแปรปรวนของประชากร จำนวน 2 ขนาด คือ 1:1 และ 1:2

5. กำหนดระดับนัยสำคัญของการทดสอบ คือ 0.05

6. ตัวสถิติทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่มที่นำมาศึกษาเปรียบเทียบกัน ประกอบด้วย t test, Mann Whitney U test, Van der Waerden test, Wald Wolfowitz Runs test และ Modified U test

7. ในแต่ละสถานการณ์จำลองชุดข้อมูลจำนวน 10,000 ชุด โดยใช้เทคนิคมอนติคาร์โล

8. เปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบพิจารณาจากความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบ โดยใช้เกณฑ์ของ Cochran (1947) ที่ระดับนัยสำคัญ 0.05

ถ้าค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 อยู่ระหว่าง 0.04 – 0.06 ถือว่าสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

ผลการวิจัย

4.1 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบ เมื่อประชากรมีการแจกแจงปกติ ผลการวิจัยดังแสดงในตารางที่ 1-2

จากตารางที่ 1 ที่ระดับนัยสำคัญ 0.05 กรณีประชากรมีการแจกแจงแบบปกติ ตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีที่สุด เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ Van der Waerden test รองลงมา คือ t test เมื่อกลุ่มตัวอย่างมีขนาดกลาง คือ t test, Mann Whitney U test, Van der Waerden test และ Modified U test และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ Van der Waerden test และ Modified U test

ตารางที่ 1 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อประชากรมีการแจกแจงปกติ

n_1	n_2	$\sigma_1^2 : \sigma_2^2$	t	M-W	V-W	W-W	M-U
5	5	1:1	0.0478*	0.0341	0.0575*	0.0150	0.0953
		1:2	0.0428*	0.0333	0.0553*	0.0156	0.0952
5	10	1:1	0.0505*	0.0435*	0.0505*	0.0048	0.0911
		1:2	0.0312	0.0341	0.0555*	0.0070	0.0814
20	20	1:1	0.0489*	0.0477*	0.0484*	0.0330	0.0581*
		1:2	0.0478*	0.0477*	0.0455*	0.0380	0.0561*
25	20	1:1	0.0470*	0.0492*	0.0499*	0.0325	0.0588*
		1:2	0.0530*	0.0511*	0.0526*	0.0393	0.0586*
50	50	1:1	0.0464*	0.0481*	0.0476*	0.0357	0.0509*
		1:2	0.0490*	0.0516*	0.0495*	0.0422*	0.0538*
50	100	1:1	0.0527*	0.0517*	0.0480*	0.0416*	0.0550*
		1:2	0.0333	0.0395	0.0585*	0.0501*	0.0524*

*สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

ตารางที่ 2 กำลังของการทดสอบ เมื่อประชากรมีการแจกแจงปกติ

n_1	n_2	$\sigma_1^2 : \sigma_2^2$	t	M-W	V-W	W-W	M-U
5	5	1:1	0.5852**	0.4810	0.5830	0.2538	0.6952
		1:2	0.4455	0.3592	0.4520**	0.1805	0.5571
5	10	1:1	0.7260**	0.6776	0.1389	0.1460	0.7840
		1:2	0.5333	0.5291	0.1234**	0.0966	0.6895
20	20	1:1	0.9966	0.9973	0.9956	0.7706	0.9977**
		1:2	0.9685	0.9753	0.9659	0.5922	0.9768**
25	20	1:1	0.9978	0.9982	0.9977	0.8041	0.9983**
		1:2	0.9788	0.9839**	0.9784	0.6623	0.9837
50	50	1:1	1.0000**	1.0000**	1.0000**	0.9812	1.0000**
		1:2	1.0000**	1.0000**	1.0000**	0.9085	1.0000**
50	100	1:1	1.0000**	1.0000**	0.9760	0.9960	1.0000**
		1:2	1.0000	1.0000	0.8905	0.9409	1.0000**

**มีกำลังของการทดสอบสูงสุดในกลุ่มที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

จากตารางที่ 2 ที่ระดับนัยสำคัญ 0.05 กรณีประชากรมีการแจกแจงปกติ ตัวสถิติทดสอบที่มีอำนาจการทดสอบสูงสุดและสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ t test และ Van der Waerden test เมื่อกลุ่มตัวอย่างมีขนาดกลางและขนาดใหญ่ คือ Modified U test

4.2 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบ เมื่อประชากรมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ (S_k =-0.65, K=6.00) ผลการวิจัยดังแสดงในตารางที่ 3-4

ตารางที่ 3 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อประชากรมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ

n_1	n_2	$\sigma_1^2 : \sigma_2^2$	t	M-W	V-W	W-W	M-U
5	5	1:1	0.0366	0.0294	0.0563*	0.0158	0.0950
		1:2	0.0409*	0.0349	0.0591*	0.0171	0.1031
5	10	1:1	0.0412*	0.0377	0.0451*	0.0046	0.0832
		1:2	0.0315	0.0340	0.0504*	0.0092	0.0820
20	20	1:1	0.0492*	0.0498*	0.0501*	0.0333	0.0595*
		1:2	0.0431*	0.0499*	0.0464*	0.0389	0.0591*
25	20	1:1	0.0469*	0.0464*	0.0472*	0.0357	0.0558*
		1:2	0.0477*	0.0479*	0.0473*	0.0338	0.0503*
50	50	1:1	0.0475*	0.0497*	0.0487*	0.0351	0.0531*
		1:2	0.0509*	0.0635	0.0586*	0.0435*	0.0653
50	100	1:1	0.0500*	0.0486*	0.0515*	0.0401*	0.0527*
		1:2	0.0326	0.0539*	0.0604	0.0519*	0.0680

*สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

ตารางที่ 4 กำลังของการทดสอบ เมื่อประชากรมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ

α	n_1	n_2	$\sigma_1^2 : \sigma_2^2$	t	M-W	V-W	W-W	M-U
0.05	5	5	1:1	0.5852	0.5015	0.5867**	0.3050	0.6855
			1:2	0.4809**	0.3981	0.4802	0.2363	0.5793
	5	10	1:1	0.7130**	0.7146	0.1587	0.1988	0.8236
			1:2	0.5439	0.5812	0.1413**	0.1602	0.7458
	20	20	1:1	0.9834	0.9984	0.9949	0.8573	0.9985**
			1:2	0.9197	0.9828**	0.9688	0.7534	0.9816
	25	20	1:1	0.9893	0.9990**	0.9976	0.8956	0.9989
			1:2	0.9387	0.9884**	0.9820	0.8256	0.9858
	50	50	1:1	1.0000**	1.0000**	1.0000**	0.9947	1.0000**
			1:2	0.9991	1.0000	1.0000**	0.9796	1.0000
	50	100	1:1	1.0000**	1.0000**	0.9711	0.9990	1.0000**
			1:2	1.0000	1.0000**	0.9004	0.9922	1.0000

**มีกำลังของการทดสอบสูงสุดในกลุ่มที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

จากตารางที่ 3 ที่ระดับนัยสำคัญ 0.05 กรณีประชากรมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ (S_k = -0.65, K=6.00) ตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีที่สุด เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ Van der Waerden test เมื่อกลุ่มตัวอย่างมีขนาดกลาง คือ t test, Mann Whitney U test, Van der Waerden test และ Modified U test และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ t test, Mann Whitney U test, Van der Waerden test และ Wald Wolfowitz Runs test

จากตารางที่ 4 ที่ระดับนัยสำคัญ 0.05 กรณีประชากรมีการแจกแจงแบบเบ้ซ้ายและความโด่งสูงกว่าปกติ (S_k = -0.65, K=6.00) ตัวสถิติทดสอบที่มีกำลังของการทดสอบสูงสุดและสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ t test และVan der Waerden test เมื่อกลุ่มตัวอย่างมีขนาดกลางและขนาดใหญ่ คือ Mann Whitney U test

4.3 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบ เมื่อประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ (S_k =0.65, K=6.00) ผลการวิจัยดังแสดงในตารางที่ 5-6

จากตารางที่ 5 ที่ระดับนัยสำคัญ 0.05 กรณีประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ (S_k = 0.65, K=6.00) ตัวสถิติทดสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีที่สุด เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ Van der Waerden test รองลงมา คือ t test เมื่อกลุ่มตัวอย่างมีขนาดกลาง คือ t test, Mann Whitney U test และ Van der Waerden test และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ t test, Mann Whitney U test, Van der Waerden test และ Wald Wolfowitz Runs test

จากตารางที่ 6 ที่ระดับนัยสำคัญ 0.05 กรณีประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ (S_k = 0.65, K=6.00) ตัวสถิติทดสอบที่มีกำลังของการทดสอบสูงสุดและสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก คือ Van der Waerden test และ t test เมื่อกลุ่มตัวอย่างมีขนาดกลาง คือ Mann Whitney U test และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ t test และ Mann Whitney U test

ตารางที่ 5 ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 เมื่อประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ

n_1	n_2	$\sigma_1^2 : \sigma_2^2$	t	M-W	V-W	W-W	M-U
5	5	1:1	0.0408*	0.0312	0.0569*	0.0150	0.0930
		1:2	0.0425*	0.0344	0.0579*	0.0167	0.0971
5	10	1:1	0.0481*	0.0394	0.0485*	0.0049	0.0893
		1:2	0.0291	0.0348	0.0527*	0.0080	0.0843
20	20	1:1	0.0477*	0.0492*	0.0477*	0.0348	0.0571*
		1:2	0.0522*	0.0571*	0.0544*	0.0388	0.0655
25	20	1:1	0.0487*	0.0488*	0.0502*	0.0323	0.0599*
		1:2	0.0517*	0.0586*	0.0584*	0.0436*	0.0649
50	50	1:1	0.0485*	0.0491*	0.0494*	0.0364	0.0525*
		1:2	0.0494*	0.0624	0.0575*	0.0414*	0.0654
50	100	1:1	0.0486*	0.0445*	0.0492*	0.0434*	0.0489*
		1:2	0.0338	0.0558*	0.0620	0.0484*	0.0709

*สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

ตารางที่ 6 กำลังของการทดสอบ เมื่อประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ

n_1	n_2	$\sigma_1^2 : \sigma_2^2$	t	M-W	V-W	W-W	M-U
5	5	1:1	0.5877	0.5005	0.5936**	0.2932	0.6931
		1:2	0.4479	0.3909	0.4808**	0.2093	0.5878
5	10	1:1	0.7105**	0.6908	0.1397	0.2302	0.7588
		1:2	0.5305	0.5708	0.1324**	0.1447	0.6938
20	20	1:1	0.9846	0.9974**	0.9943	0.8536	0.9974**
		1:2	0.9437	0.9837**	0.9697	0.6697	0.9851
25	20	1:1	0.9904	0.9989**	0.9976	0.8784	0.9989**
		1:2	0.9686	0.9914**	0.9850	0.7202	0.9910
50	50	1:1	1.0000**	1.0000**	1.0000**	0.9956	1.0000**
		1:2	1.0000**	1.0000	1.0000**	0.9482	1.0000
50	100	1:1	1.0000**	1.0000**	0.9813	0.9994	1.0000**
		1:2	1.0000	1.0000**	0.9201	0.9828	1.0000

**มีกำลังของการทดสอบสูงสุดในกลุ่มที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

สรุปผลการวิจัย

สรุปตัวสถิติทดสอบที่เหมาะสมในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่ม โดยพิจารณาจากความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และกำลังของการทดสอบ แสดงดังตารางที่ 7

ตารางที่ 7 ตัวสถิติทดสอบที่เหมาะสมแต่ละสถานการณ์ในการทดสอบความแตกต่างของค่ากลางระหว่างประชากร 2 กลุ่ม

การแจกแจง	ขนาดตัวอย่าง		
	เล็ก	กลาง	ใหญ่
ปกติ	t test, Van der Waerden test	Modified U test	Modified U test
เบ้ซ้ายและความโด่งสูงกว่าปกติ	t test, Van der Waerden test	Mann Whitney U test	Mann Whitney U test
เบ้ขวาและความโด่งสูงกว่าปกติ	t test, Van der Waerden test	Mann Whitney U test	t test, Mann Whitney U test

จากตารางที่ 7 แสดงให้เห็นว่า เมื่อประชากรมีการแจกแจงปกติ กลุ่มตัวอย่างมีขนาดเล็ก ตัวสถิติทดสอบที่เหมาะสม คือ t test และ Van der Waerden test กลุ่มตัวอย่างมีขนาดกลางและขนาดใหญ่ คือ Modified U test เมื่อประชากรมีการแจกแจงเบ้ซ้ายและความโด่งสูงกว่าปกติ กลุ่มตัวอย่างมีขนาดเล็ก ตัวสถิติทดสอบที่เหมาะสม คือ t test และVan der Waerden test กลุ่มตัวอย่างมีขนาดกลางและขนาดใหญ่ คือ Mann Whitney U test และเมื่อประชากรมีการแจกแจงเบ้ขวาและความโด่งสูงกว่าปกติ กลุ่มตัวอย่างมีขนาดเล็ก ตัวสถิติทดสอบที่เหมาะสม คือ t test และVan der Waerden test กลุ่มตัวอย่างมีขนาดกลาง คือ Mann Whitney U test และกลุ่มตัวอย่างมีขนาดใหญ่ คือ t test และ Mann Whitney U test

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิที่ให้ทุนอุดหนุนการวิจัย และขอขอบคุณคณะกรรมการผู้ทรงคุณวุฒิที่กรุณาตรวจสอบให้ข้อเสนอแนะในการปรับปรุงต้นฉบับบทความวิจัยดังกล่าวนี้

เอกสารอ้างอิง

ศิริชัย กาญจนวาสี. (2545). สถิติประยุกต์สำหรับการวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Boneau, C.A. (1960). The effects of violation of assumptions underlying the t-test. Psychological Bulletin 57: 49-64.

Cochran, W.G. (1947). Some consequences when The assumptions for the analysis of variance are not satisfied. Biometrics 3: 22-38.

Fligner, M. A., and Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem. Journal of the American Statistical Association 81.

Hopkins, K.D., Glass, G.V., and Hopkins, B.R. (1987). Basic statistics for the behavioral sciences. 2nd ed. Prentice-Hall: Englewood Cliffs, NJ.

Hsu, P.L. (1938). Contributions to the theory of Student's t test as applied to the problem of two samples. Statistical Research Memoirs 2: 1-24.

Scheffe, H. (1959). The analysis of variance. New York: Wiley.

Scheffe, H. (1970). Practical solutions of the Behrens-Fisher problem. Journal of the American Statistical Association 65: 1501-1508.

Ramberg, J.S., Dudewicz, E.J., Tadikamalla, P.R., and Mykytka, E.F. (1979). A probability distribution and its uses in fitting data. Technometrics 21: 201-214.

Sheskin, J.D. (2000). Handbook of parametric and nonparametric statistical procedures. 2nd ed. Florida: Chapman and Hall/CRC.

Van der Waerden, B.L. (1952). Order tests for the two-sample problem and their power. Indagationes Mathematicae. 14: 453–458.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. Ann. Math Statist 11: 147-162.

Zimmerman, D.W., and Zumbo, B.D. (2009). Hazards in Choosing Between Pooled and Separate-Variances t Tests. Psicologica 30: 371-390.

❑❑❑❑❑