บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การกระจายแบบที (t-distribution)

W.S. Gosset เป็นผู้ค้นพบการกระจายแบบที หรือ Student's t-distribution ในปี ค.ศ. 1908 เนื่องจากในขณะนั้นนายจ้างห้ามมิให้เขาเขียนหรือจัดพิมพ์เอกสารใด ๆ เขาจึงแอบพิมพ์เผยแพร่โดยใช้นามปากกาว่า "Student" ที่มาของการกระจายแบบทีเกี่ยวข้องกับการกระจายแบบโค้งปกติและไค-สแคว์ กล่าวคือ ถ้ามีกลุ่มตัวอย่างซึ่งมาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบปกติมีมัชฌิมเลขคณิตเท่ากับ 0 ความแปรปรวนเท่ากับ 1 อยู่กลุ่มหนึ่ง สุ่มตัวอย่างมา 1 ตัวเพื่อหาค่าซี (Z) หลังจากนั้นสุ่มตัวอย่างมาอีก 10 ตัวเพื่อหาค่า χ^2_{10} อัตราส่วนของค่าซีและ χ^2_{10} คือ t_{10} นั่นคือ

$$t_{10} = \frac{Z}{\sqrt{\frac{\chi^2_{10}}{10}}}$$

และรูปทั่วไปของการกระจายแบบที คือ

$$t_n = \frac{Z}{\sqrt{\frac{\chi^2_n}{n}}}$$ เมื่อ n คือ degrees of freedom

<u>คุณสมบัติของการกระจายแบบที</u>

1. มีลักษณะสมมาตร
2. ค่าเฉลี่ย มัธยฐาน และฐานนิยมมีค่าเท่ากับ 0
3. ความแปรปรวนเท่ากับ $\frac{n}{n-2}$ (เมื่อ n เป็น degrees of freedom)

ส่วนเบี่ยงเบนมาตรฐานจะมีค่ามากกว่า 1 เล็กน้อย ในกรณีที่ n มีค่ามาก ๆ ความแปรปรวนจะมีค่าเท่ากับ 1

4. ถ้า n มีค่ามาก ๆ การกระจายจะมีลักษณะใกล้เคียงหรือเป็นแบบปกติ

5. ${}_{\alpha}t_n = (1 - \alpha)\ t_n$

การทดสอบที (t-test)

W.S. Gosset เป็นผู้สร้างสถิติทดสอบนี้ขึ้นมาเมื่อ ค.ศ. 1908 พร้อมกับการกระจายแบบที และ R.A. Fisher เป็นผู้เน้นบทบาทของสถิติทดสอบนี้

ข้อตกลงเบื้องต้น

1. ลักษณะการแจกแจงของประชากรเป็นแบบปกติ

2. กลุ่มตัวอย่างทั้งสองเป็นอิสระต่อกัน (independent groups) และเป็นกลุ่มตัวอย่างสุ่ม

3. ตัวอย่างในกลุ่มเป็นอิสระต่อกัน (independent within samples)

4. ข้อมูลจัดอยู่ในสเกลวัดได้ (measurable scale)

5. ความแปรปรวนของประชากรเท่ากัน (Homogeneity of variance)

การคำนวณค่าทดสอบ

ถ้าให้ $X_1, X_2, \ldots, X_m$ และ $Y_1, Y_2, \ldots, Y_n$ เป็นกลุ่มตัวอย่างสุ่มจากประชากรสองกลุ่มซึ่งมีลักษณะการแจกแจงเป็นแบบปกติ จะคำนวณค่าทีได้จากสูตร

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\Sigma x^2 + \Sigma y^2}{(m+n-2)}\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}$$

โดยที่ $\Sigma x^2 = \Sigma(X-\overline{X})^2$

$\Sigma y^2 = \Sigma(Y-\overline{Y})^2$

ค่า t จะกระจายแบบทีด้วย m+n-2 degree of freedom

การทดสอบของวิลค๊อกซอน

(Wilcoxon Test or Wilcoxon's Rank - Sum Test)

วิลค๊อกซอนเป็นผู้สร้างสถิติทดสอบนี้ขึ้นเมื่อ ค.ศ. 1949 เพื่อศึกษาการเปรียบเทียบความแตกต่างอย่างมีนัยสำคัญทางสถิติของประชากรซึ่งเป็นที่มาของกลุ่มตัวอย่างสุ่มสองกลุ่มที่เป็นอิสระต่อกัน สถิติทดสอบของวิลค๊อกซอนมีพื้นฐานการสร้างเช่นเดียวกับการทดสอบของฟิชเชอร์ (Fisher randomization t-test) แต่มีความไว (sensitive) มากกว่า ความแตกต่างอยู่ที่ว่า วิธีการของวิลค๊อกซอนกลุ่มตัวอย่างจะถูกแทนที่โดยใช้อันดับที่ได้มาจากกลุ่มตัวอย่างทั้งหมด และจะทดสอบข้อมูลโดยใช้ค่าอันดับที่ได้นี้ แต่วิธีการของฟิชเชอร์จะใช้ค่าที่ได้จากการสังเกตมาคิดคำนวณเพื่อทดสอบข้อมูล วิลค๊อกซอนมีเหตุผลว่า เมื่อมีกลุ่มตัวอย่าง n คนจากประชากรกลุ่มหนึ่ง และกลุ่มตัวอย่าง m คนจากประชากรอีกกลุ่มหนึ่ง โดยมีสมมุติฐานที่ต้องการทดสอบคือ ประชากรซึ่งเป็นที่มาของกลุ่มตัวอย่างทั้งสองไม่แตกต่างกัน การนำค่าอันดับทั้งหมดมาเลือกครั้งละ n ตัวแล้วหาผลรวม จะสามารถหาผลรวมได้ทั้งหมด $\binom{n+m}{n}$ จำนวนซึ่งจะมีการกระจายในลักษณะสุ่ม และถ้าสมมุติฐานสูญเป็นจริงก็จะต้องเกิดการกระจายในลักษณะสุ่มขึ้นจริง ด้วยเหตุนี้สมมุติฐานสูญจะถูกปฏิเสธถ้าการกระจายของผลรวมของอันดับไม่เป็นการกระจายแบบสุ่ม นั่นคือสมมุติฐานสูญจะถูกปฏิเสธถ้าค่าผลรวมของอันดับของกลุ่มตัวอย่าง n คนตกอยู่ในเขตวิกฤตซึ่งจะมีค่าเท่ากับ $\alpha\binom{n+m}{n}$ ค่าสำหรับการทดสอบทางเดียวและเท่ากับ $\alpha\binom{n+m}{n}/2$ สำหรับการทดสอบสองทาง

<u>ข้อตกลงเบื้องต้น</u> (Bradley 1964: 107)

1. กลุ่มตัวอย่างทั้งสองเป็นอิสระต่อกัน (independent groups) และเป็นกลุ่มตัวอย่างสุ่ม

2. ตัวอย่างในกลุ่มเป็นอิสระต่อกัน (independent within samples)

3. ไม่มีอันดับที่ซ้ำกัน (tied rank)

<u>การคำนวณค่าทดสอบ</u>

เอาคะแนนที่ได้จากการสังเกตทั้งหมดมาจัดอันดับแล้วรวมค่าอันดับของแต่ละกลุ่ม ให้ผลรวมของค่าอันดับที่น้อยกว่าเป็น T1 และผลรวมอีกค่าหนึ่งเป็น T2 ค่า T1 ที่ได้นี้จะเป็นค่าที่นำไปเปรียบเทียบกับเขตวิกฤตซึ่งมีวิธีการคำนวณดังนี้คือ นำค่าอันดับทั้งหมดมาเลือกครั้งละ n ตัว (n คือจำนวนตัวอย่างของกลุ่มตัวอย่างที่มีค่าผลรวมของค่าอันดับเป็น T1) แล้วหาผลรวม จะได้ค่าผลรวมทั้งหมด $\binom{n+m}{n}$ ค่า เรียงลำดับค่าผลรวมที่ได้จากน้อยที่สุดไปหามากที่สุด ตัดค่าผลรวมที่เรียงลำดับแล้ว $\alpha\binom{n+m}{n}$ ค่าสำหรับการทดสอบหางเดียว หรือ $\alpha\binom{n+m}{n}/2$ ค่าสำหรับการทดสอบสองหางเป็นเขตวิกฤต ถ้าค่า T1 ตกอยู่ในเขตวิกฤตก็จะปฏิเสธสมมุติฐานสูญ (Ho)

<u>การทดสอบของวิลค๊อกซอน เมื่อกลุ่มตัวอย่างมีขนาดใหญ่</u>

ถ้าจำนวนตัวอย่างในแต่ละกลุ่มมากกว่า 10 จำนวนขึ้นไป ถือว่าเป็นกลุ่มตัวอย่างที่มีขนาดใหญ่ ซึ่งสามารถคำนวณหาค่าของซีได้ดังสูตร

$$Z = \frac{T1 - E(T1)}{\sqrt{Var\ (T1)}} = \frac{T1 - n(n+m+1)/2}{\sqrt{(nm/12)\ (n+m+1)}}$$

เมื่อ n คือ จำนวนตัวอย่างของกลุ่มตัวอย่างที่มีผลรวมของค่าอันดับเป็น T1

m คือ จำนวนตัวอย่างของกลุ่มตัวอย่างที่มีผลรวมของค่าอันดับเป็น T2

T1 คือ ผลรวมของค่าอันดับที่น้อยกว่า

T2 คือ ผลรวมของค่าอันดับที่มากกว่า

Z จะแจกแจงเป็น N(0,1) ซึ่งค่าวิกฤตสามารถหาได้จากตารางพื้นที่ใต้โค้งปกติ

นอร์มอล-สกอร์เทส

(Normal-Scores Test)

อาศัยวิธีการของฟิชเชอร์ (Fisher randomization t-test) นอร์มอลสกอร์เกิดจากการสร้างจุดกึ่งกลางของค่าที่ได้มาจากการสังเกตซึ่งตามสภาพความเป็นจริงแล้วไม่มีจุดนี้ วิธีการนี้จะทำให้มีอำนาจของการทดสอบสูงขึ้น และ A.R.E. เมื่อเทียบกับสถิติพาราមิเตอร์อื่นซึ่งสามารถใช้แทนกันได้มีค่าเท่ากับ 1 เมื่อข้อมูลเป็นไปตามข้อตกลงเบื้องต้นของพารามิเตอร์นั้น ในกรณีที่ต้องการใช้สถิติทดสอบประเภทแรงค์แรนดอมไมเซชัน (rank-randomization) ทำการทดสอบข้อมูลจำเป็นต้องแปลงคะแนนที่ได้จากการสังเกตมาเป็นอันดับ แต่วิธีการนี้ทำให้สูญเสียค่า A.R.E. ไปบ้างเล็กน้อยซึ่งจะมีผลทำให้กราฟของ A.R.E. ไม่ถึง 1 เพียงแต่เข้าใกล้เท่านั้น นอร์มอลสกอร์สร้างขึ้นมาเพื่อช่วยปรับให้ค่า A.R.E. สูงถึง 1 ซึ่งมีวิธีการดังต่อไปนี้ แปลงคะแนนที่ได้จากการสังเกตมาเป็นอันดับ แปลงคะแนนอีกครั้งหนึ่งโดยการแปลงค่าอันดับให้เป็นค่าคาดหวังที่สัมพันธ์โดยตรงกับ normal order statistics วิธีการของนอร์มอล-สกอร์เทสต้องใช้ตารางถึง 2 ตารางด้วยกัน คือ

1. ตารางนอร์มอลสกอร์
2. ตารางทีหรือตารางซี

นอร์มอล-สกอร์เทส มีหลายวิธีด้วยกันซึ่งจะมีคะแนนนอร์มอลแตกต่างกันตามชนิดของการทดสอบ แต่มีหลักฐานและวิธีการในการแปลงอันดับให้เป็นคะแนนนอร์มอลคล้าย ๆ กัน การแปลงคะแนนให้เป็นคะแนนนอร์มอลไม่มีวิธีการอื่นที่จะใช้แปลงได้นอกจากอาศัยตารางนอร์มอลสกอร์ ถึงแม้ว่านอร์มอล-สกอร์เทส เป็นสถิติทดสอบที่ยุ่งยากมากกว่าสถิติทดสอบแบบอื่น ๆ ก็ตาม แต่ถ้าพิจารณาทางด้านประสิทธิภาพแล้วนอร์มอล-สกอร์เทสมีประสิทธิภาพสูงกว่าสถิติทดสอบพาราเมตริกและนันพาราเมตริกแบบอื่น ๆ

การทดสอบของ เทอรี่-โฮฟฟ์ดิง

(Terry-Hoeffding Normal-Scores Test)

เทอรี่และโฮฟฟ์ดิงเป็นผู้สร้างสถิติทดสอบนี้ขึ้นมา เมื่อ ค.ศ. 1952 เขาสมมุติว่าคะแนนที่ได้มาจากการสังเกตไม่มีคะแนนกึ่งกลางและจะใช้ค่า $E(V^{(i)})$ (The expected normal order statistic or The expected normal scores) มาสร้างให้เกิดคะแนนกึ่งกลางขึ้นใหม่ โดยค่า $E(V^{(i)})$ นี้จะสัมพันธ์โดยตรงกับค่าอันดับของคะแนนแต่ละคะแนน หรือกล่าวอีกอย่างหนึ่งว่า เราจะแบ่งครึ่งคะแนนที่ได้มาจากการสังเกตโดยใช้ค่า $E(V^{(i)})$ เป็นตัวแบ่ง

การแปลงคะแนนให้เป็น The Expected Normal Scores

การแปลงคะแนนให้เป็น The Expected Normal Scores นั้นจะต้องจัดอันดับคะแนนที่ได้มาทั้งหมด ฉะนั้นคะแนนที่ได้จากการสังเกตทุกค่าจะมีค่าอันดับประจำอยู่ซึ่งให้เป็น $V^{(i)}$ ซึ่งเป็นคะแนนอันดับในตำแหน่งที่ i แบ่งครึ่งคะแนนอันดับและให้ส่วนที่มีคะแนนอันดับต่ำกว่ามีค่าเป็นลบ ส่วนที่มีอันดับสูงกว่ามีค่าเป็นบวก นำค่าอันดับที่ได้ไปเปิดตาราง Expected values of order statistic of the Terry-Hoeffding form $E(V^{(i)})$ (Marascuilo 1977: 509-510) เพื่อหาค่าคาดหวังของคะแนนในตำแหน่งที่ i $[E(V^{(i)})]$ ซึ่งอันดับที่เข้าใกล้กลาง ๆ จะมีค่า $E(V^{(i)})$ เข้าใกล้ศูนย์และผลรวมของ $E(V^{(i)})$ ทั้งหมดจะมีค่าเท่ากับศูนย์ Fisher และ Yates (Bradley 1968: 148) ได้กล่าวไว้ในตำราสถิติพื้นฐานซึ่งเกี่ยวกับตารางสถิติว่า ในการแปลงค่าอันดับให้เป็นคะแนนนอร์มอลนั้นเปรียบเสมือนการคิดว่าอันดับทั้งหมดเป็นเลขสุ่มที่ได้มาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบปกติมีมัชฌิมเลขคณิตเท่ากับ o และความแปรปรวนเท่ากับ 1 Fisher และ Yates ได้สร้างตารางคะแนนนอร์มอลซึ่งให้คะแนนนอร์มอลเป็นเลขทศนิยม และตารางของเขาแพร่หลายมากโดย Harter

ข้อตกลงเบื้องต้น (Bradley 1964: 150)

1. กลุ่มตัวอย่างทั้งสองเป็นอิสระต่อกัน (independent groups) และเป็นกลุ่มตัวอย่างสุ่ม

2. ตัวอย่างในกลุ่มเป็นอิสระต่อกัน (independent within samples)

3. ไม่มีอันดับที่ซ้ำกัน (tied rank)

<u>การคำนวณค่าทดสอบ</u>

นำคะแนนที่ได้มาจากการสังเกตทั้งหมดมาจัดอันดับ นำค่า N (จำนวนตัวอย่างทั้งหมด) และค่าอันดับมาหาค่า $E(V^{(i)})$ โดยการเปิดตาราง Expected Values of order statistic of the Terry-Hoeffding form $E(V^{(i)})$ เมื่อได้ค่าของ $E(V^{(i)})$ แล้วรวมค่า $E(V^{(i)})$ ของแต่ละกลุ่ม ให้ผลรวมของ $E(V^{(i)})$ ที่มีค่าน้อยกว่าเป็น T1 และอีกค่าหนึ่งเป็น T2 ค่า T1 ที่ได้นี้จะเป็นค่าที่นำไปเปรียบเทียบกับเขตวิกฤตซึ่งมีวิธีการคำนวณดังนี้คือ นำค่า $E(V^{(i)})$ ทั้งหมดมาเลือกครั้งละ n ตัว (n คือจำนวนตัวอย่างของกลุ่มตัวอย่างที่มีค่าผลรวมของค่า $E(V^{(i)})$ เป็น T1) แล้วหาผลรวมจะได้ค่าผลรวมทั้งหมด $\binom{n+m}{n}$ ค่า เรียงลำดับค่าผลรวมที่ได้จากน้อยที่สุดไปหามากที่สุด ตัดค่าผลรวมที่เรียงลำดับแล้ว $\alpha\binom{n+m}{n}$ ค่าสำหรับการทดสอบหางเดียวหรือ $\alpha\binom{n+m}{n}/2$ ค่าสำหรับการทดสอบสองหางเป็นเขตวิกฤต ถ้าค่า T1 ตกอยู่ในเขตวิกฤตก็จะปฏิเสธสมมุติฐานสูญ (Ho)

<u>การทดสอบของเทอรี่-โฮฟฟ์ดิงเมื่อกลุ่มตัวอย่างมีขนาดใหญ่</u>

ถ้าจำนวนตัวอย่างในแต่ละกลุ่มมากกว่า 7 จำนวนขึ้นไปถือว่าเป็นกลุ่มตัวอย่างที่มีขนาดใหญ่ ซึ่งสามารถคำนวณหาค่าของซีได้ดังสูตร

$$Z = \frac{T1 - E(T1)}{\sqrt{Var(T1)}}$$

$$T1 = \sum_{i=1}^{n} E(V^{(i)})$$

$$Var(T1) = \frac{nm}{N-1} \frac{\sum_{i=1}^{n} E(V^{(i)})^2}{N}$$

$E(T1) = nE[E(V^{(i)})] = 0$ เพราะว่า $E[E(V^{(i)})] = \frac{1}{N}\sum_{i=1}^{n} E(V^{(i)}) = 0$

Z จะแจกแจงเป็น N(0,1) ซึ่งค่าวิกฤตสามารถหาได้จากตารางพื้นที่ใต้โค้งปกติ

การทดสอบของแวน เดอ แวร์เดน

(Van der Waerden Normal-Scores Test)

แวน เดอ แวร์เดน เป็นผู้สร้างสถิติทดสอบนี้ขึ้นมาเมื่อ ค.ศ. 1953 โดยใช้ค่าของ inverse-normal scores ($Z^{(i)}$) แทนค่าที่ได้จากการสังเกต

การแปลงคะแนนให้เป็น inverse-normal scores ($Z^{(i)}$)

ให้ $r_1, r_2, r_3, \ldots, r_n$ เป็นค่าอันดับของคะแนนที่ได้มาจากการสังเกต และให้ p_i เป็นเปอร์เซ็นไทล์แรงค์ (percentile rank) ซึ่งสัมพันธ์กับ normalized observation จะได้ว่า

$$p_i = \frac{r_i}{N+1} \quad , \quad i = 1, 2, 3, \ldots, n \qquad N = n+m$$

ถ้าให้ $\phi(Z^{(i)})$ คือ comulative distribution ของ $Z^{(i)}$ จะได้ว่า

$$p_i = \phi(Z^{(i)})$$

$$Z^{(i)} = \phi^{-1}(p_i)$$

$$= \phi^{-1}\left(\frac{r_i}{N+1}\right)$$

ซึ่งเราอาจจะคำนวณค่าของ $Z^{(i)}$ ได้โดยใช้สูตรนี้ หรือใช้เปิดค่า $Z^{(i)}$ จากตารางที่ได้คำนวณค่า $Z^{(i)}$ เมื่อทราบค่า p_i (Marascuilo 1977: 485) **หรือเมื่อทราบค่าของ** r_i (Marascuilo 1977: 511) ไว้เรียบร้อยแล้ว

<u>ข้อตกลงเบื้องต้น</u>

1. กลุ่มตัวอย่างทั้งสองเป็นอิสระต่อกัน (independent groups) และเป็นกลุ่มตัวอย่างสุ่ม

2. ตัวอย่างในกลุ่มเป็นอิสระต่อกัน (independent within samples)

3. ไม่มีอันดับที่ซ้ำกัน (tied rank)

<u>การคำนวณค่าทดสอบ</u>

เอาคะแนนที่ได้มาจากการสังเกตทั้งหมดมาจัดอันดับ นำค่า N และค่าอันดับมาหาค่า $Z^{(i)}$ เมื่อได้ค่า $Z^{(i)}$ แล้วรวมค่า $Z^{(i)}$ ของแต่ละกลุ่ม ให้ผลรวมของ $Z^{(i)}$ ที่มีค่าน้อยว่าเป็น T1 และอีกค่าหนึ่งเป็น T2 ค่า T1 ที่ได้นี้จะเป็นค่าที่นำไปเปรียบเทียบกับเขตวิกฤตซึ่งมีวิธีการคำนวณดังนี้ นำค่า $Z^{(i)}$ ทั้งหมดมาเลือกครั้งละ n ตัว แล้วหาผลรวม จะได้ค่าผลรวมทั้งหมด $\binom{n+m}{n}$ ค่า เรียงลำดับค่าผลรวมที่ได้จากน้อยที่สุดไปหามากที่สุด ตัดค่าผลรวมที่เรียงลำดับแล้ว $\alpha\binom{n+m}{n}$ ค่าสำหรับการทดสอบหางเดียวหรือ $\alpha\binom{n+m}{n}/2$ ค่าสำหรับการทดสอบสองหาง ถ้าค่า T1 ตกอยู่ในเขตวิกฤตก็จะปฏิเสธสมมุติฐานสูญ (Ho)

<u>การทดสอบของแวน เดอ แวร์เดน เมื่อกลุ่มตัวอย่างมีขนาดใหญ่</u>

ถ้าจำนวนตัวอย่างในแต่ละกลุ่มมากกว่า 7 จำนวนขึ้นไปถือว่าเป็นกลุ่มตัวอย่างที่มีขนาดใหญ่ ซึ่งสามารถคำนวณหาค่าของซีได้ดังสูตร

$$Z = \frac{T1 - E(T1)}{\sqrt{Var(T1)}}$$

เมื่อ $$T1 = \sum_{i=1}^{n} Z^{(i)}$$

$$Var(T1) = \frac{nm}{N-1}\sum_{i=1}^{n}\frac{(Z^{(i)})^2}{N}$$

$$E(T1) = 0$$

Z จะแจกแจงเป็น N(0,1) ซึ่งค่าวิกฤตสามารถหาได้จากตารางพื้นที่ใต้โค้งปกติ

อนึ่งมีข้อสังเกตเกี่ยวกับการใช้การทดสอบทั้ง 4 วิธีดังนี้คือ ในกรณีที่ข้อมูลที่ใช้ในการทดสอบอยู่ในมาตราอันตรภาค (interval scale) หรือมาตราอัตราส่วน (ratio scale) สามารถใช้การทดสอบทั้ง 4 วิธีทำการทดสอบได้ แต่ถ้าข้อมูลที่ใช้ในการทดสอบอยู่ในรูปของอันดับ (rank) จะใช้การทดสอบทีทำการทดสอบไม่ได้ สถิติทดสอบที่จะใช้ทดสอบข้อมูลที่อยู่ในรูปของอันดับได้ คือ การทดสอบของวิลค๊อกซอน, การทดสอบของเทอรี่-โฮฟฟ์ดิง และการทดสอบของแวน เดอ แวร์เดน

งานวิจัยที่เกี่ยวข้อง

การศึกษาทฤษฎีการสุ่มตัวอย่าง การคำนวณค่าทางสถิติ และทฤษฎีของ asymtotic ได้รับการนำมาใช้ในการพิสูจน์ความจริงที่ว่า การทดสอบทีและการทดสอบของวิลค๊อกซอนมีอำนาจของการทดสอบเกือบจะเท่าๆ กันเมื่อกลุ่มตัวอย่างสุ่มทั้งสองกลุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบปกติ (Dixon, 1954; Hodges & Lehmann, 1956; Lehmann, 1975; Neave & Granger, 1968, cited by Blair & Higgins 1980: 311) ผลการวิจัยปรากฏว่า การทดสอบทีมีอำนาจของการทดสอบมากกว่าการทดสอบของวิลค๊อกซอน

Hodges และ Lehmann (1956, cited by Blair & Higgins 1980: 311) ได้ทำการศึกษาจุดที่น่าสนใจและศักยภาพที่สำคัญของ asymtotic เขาได้พิสูจน์ว่า ค่า A.R.E. (หรือ Pitman efficiency) ของการทดสอบของวิลค๊อกซอนซึ่งสัมพันธ์กับการทดสอบทีมีค่าสูงถึงอินฟินิตี้และมีค่าไม่ต่ำกว่า 0.864 จากผลที่ได้นี้ Hodges และ Lehmann (1956: 356) ได้กล่าวว่า

จากการทดสอบที่แล้วมาทำให้ทราบถึงประสิทธิภาพและผลที่ได้จากภาวะกลุ่มตัวอย่างที่ใช้มีขนาดต่าง ๆ กัน และแนวทางต่าง ๆ ที่จะเป็นไปได้ในทางภาคปฏิบัติทั้งหมด แสดงให้เห็นว่าการที่จะใช้การทดสอบของวิลค๊อกซอนแทนการทดสอบทีจะไม่ทำให้เกิดการสูญเสียประสิทธิภาพแต่อย่างไร (และในทางตรงกันข้ามการทดสอบของวิลค๊อกซอนอาจจะมีประสิทธิภาพในการทดสอบสูงกว่าการทดสอบทีเป็นอย่างมาก) แต่อย่างไรก็ดีการทดสอบด้วยค่า A.R.E.

เป็นการคำนวณภายใต้ข้อตกลงเบื้องต้นที่ไม่เป็นจริง ดังที่ Bradley (1968: 58) กล่าวไว้ว่า "ไม่มีการทดสอบใดที่ใช้กลุ่มตัวอย่างที่มีขนาดใหญ่มาก และนอกจากนั้นไม่มีใครสนใจอำนาจของการทดสอบที่จะปฏิเสธสมมุติฐานซึ่งแตกต่างจากสมมุติฐานสูญแต่เพียงเล็กน้อย"

Boneau (1962, cited by Blair & Higgins 1980: 312) ให้ใช้เทคนิคมอนติคาร์โลศึกษาเปรียบเทียบอำนาจของการทดสอบทีกับการทดสอบของวิลค๊อกซอน เมื่อกลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็น normal, rectangular และ exponential ได้ข้อสรุปว่า โดยทั่วไปแล้วการทดสอบทีมีอำนาจของการทดสอบสูงกว่าการทดสอบ Mann-Witney U Test (Wilcoxon Test) แต่ไม่มากนัก นอกจากนั้นเขายังมีแนวความคิดว่า ผลการศึกษาในด้าน asymtotic ของ Hodges และ Lehmann ไม่สามารถทำให้ครอบคลุมทุกสภาวะที่กลุ่มตัวอย่างมีขนาดจำกัด

Toothaker (1972, cited by Blair & Higgins 1980: 312) ให้ใช้คอมพิวเตอร์ซิมูเลชั่นเพื่อเปรียบเทียบอำนาจของการทดสอบทีกับอำนาจของการทดสอบของวิลค๊อกซอน เมื่อกลุ่มตัวอย่างทั้งสองกลุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบปกติ เป็นแบบยูนิฟอร์ม และประชากรที่มีลักษณะเบ้ (skewed populations) โดยที่กลุ่มตัวอย่างที่ใช้มีขนาดไม่เกิน 5 ผลที่ได้เป็นไปตามลักษณะเดียวกันกับผลการศึกษาของ Boneau (1962) นั่นคือการทดสอบทีมีอำนาจของการทดสอบสูงกว่าการทดสอบของวิลค๊อกซอนแต่ไม่มากนัก

Sawat Pratoomraj (1970: abstract) ได้ทำการวิจัยและพบว่า ถ้ากลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบเดียวกันแล้ว การทดสอบทีและการทดสอบของแมน-วิทนี (เป็นรูปแบบหนึ่งของการทดสอบของวิลค๊อกซอน ซึ่งปรับปรุงเป็นอิสระในปี ค.ศ. 1947) มีอำนาจของการทดสอบสูงที่สุดในประเภทของการทดสอบทั้ง 5 คือ การทดสอบที การทดสอบของแมน-วิทนี (Mann-Witney U Test) การทดสอบของเวลช์ (Welch Test) และการทดสอบซี (z-test) เมื่อความแปรปรวนของประชากรเท่ากัน

Neave และ Granger (1968, cited by Blair & Higgins 1980: 312) ได้ทำการศึกษาเปรียบเทียบอำนาจของการทดสอบทีกับการทดสอบของวิลค๊อกซอนเมื่อกลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบปกติแบบ Super Position และกลุ่มตัวอย่างมีขนาด $n_1 = n_2 = 20$ และ $n_1 = 20$, $n_2 = 40$ ผลปรากฏว่า

การทดสอบของวิลค๊อกซอนมีอำนาจของการทดสอบเหนือกว่าการทดสอบที โดยที่การทดสอบของวิลค๊อกซอนมีสัดส่วนของการเกิดนัยสำคัญมากกว่าการทดสอบที และผลต่างของสัดส่วนของการเกิดนัยสำคัญของสถิติทดสอบทั้งสองสูงถึง 0.12

Blair, Higginsและ Smitley (1980, cited by Blair & Higgins 1980: 312)ได้ทำการซิมูเลท (Simulate) โดยใช้คอมพิวเตอร์เพื่อศึกษาเปรียบเทียบอำนาจของการทดสอบทีกับการทดสอบของวิลค๊อกซอนเมื่อประชากรซึ่งเป็นที่มาของกลุ่มตัวอย่างมีลักษณะการแจกแจงเป็น exponential เขาได้สรุปว่าการทดสอบของวิลค๊อกซอนมีอำนาจของการทดสอบเหนือกว่าการทดสอบที และการที่ Boneau (1962) ศึกษาโดยใช้กลุ่มตัวอย่างเล็กทำให้ข้อสรุปของเขาผิดไปได้

Blair and Higgins(1980: 309-335)ได้ใช้เทคนิคมอนติคาร์โลทำการศึกษาเพื่อเปรียบเทียบอำนาจของการทดสอบของวิลค๊อกซอนกับการทดสอบทีภายใต้ลักษณะการแจกแจงของประชากรแบบ uniform, แบบ Laplace, แบบ half normal, แบบ exponential, แบบ mixed normal และแบบ mixed-uniform และกลุ่มตัวอย่างที่ใช้ในการศึกษามีขนาด (n_1, n_2) เท่ากับ (3,9), (6,6), (9,27), (18,18), (27,81) และ (54,54) ผลการศึกษาสรุปได้ว่า

1. สรุปโดยทั่วไปการทดสอบของวิลค๊อกซอนมีอำนาจของการทดสอบสูงกว่าการทดสอบทีมาก

2. A.R.E. เป็นดัชนีที่ดีในการชี้ถึงอำนาจของการทดสอบทั้งสอง

3. ผลการศึกษาจากกลุ่มตัวอย่างที่มีขนาดเล็กจะได้ผลที่แตกต่างไปจากการศึกษาในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่

4. เนื่องจากการศึกษาที่มีมาก่อนนี้ศึกษากับประชากรที่มีการกระจายน้อยและใช้กลุ่มตัวอย่างที่แตกต่างกันออกไปมาก ทำให้ผลที่ได้มาอยู่ในภาวะที่น่าสงสัย

Bradley (1978: 108) ได้ทำการศึกษาและพบว่า การทดสอบของวิลค๊อกซอน, การทดสอบของเทอรี่และการทดสอบของแวน เดอ แวร์เดน มีประสิทธิภาพและมีอำนาจของการทดสอบสูงกว่าการทดสอบทีเมื่อข้อมูลที่ได้มาอยู่ในรูปของอันดับ