บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในส่วนของวรรณคดีที่เกี่ยวข้อง นำเสนอเป็น 3 ตอน คือ

ตอนที่ 1 มาตราการวัด

ตอนที่ 2 สถิติทดสอบ

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

<u>ตอนที่ 1</u> มาตราการวัด (Measurement Scales)

ในการใช้สถิติเพื่อวิเคราะห์ข้อมูลในการวิจัยนั้น ผู้วิจัยจำเป็นต้องทราบลักษณะของข้อมูลที่ใช้ในการวิจัยเสียก่อน เพื่อเป็นเครื่องวินิจฉัยว่าจะใช้วิธีการทางสถิติแบบใดในการวิเคราะห์จึงจะถูกต้องเหมาะสม ดังนั้นผู้วิจัยจึงควรทราบถึงมาตราการวัดของข้อมูลโดย Stevens (1946) ได้จำแนกมาตราการวัดออกเป็น 4 ระดับ ดังนี้

1. มาตรานามบัญญัติ (Nominal Scales) เป็นมาตราการวัดในระดับต่ำสุด เป็นการวัดแบบง่าย ๆ โดยการจำแนกหรือแยกประเภท (Classification) ตามคุณลักษณะที่ไม่เหมือนกัน อาจเป็นเพียงการเรียกชื่อ ดังนั้นการวัดในระดับนี้บางทีจึงไม่เป็นที่ยอมรับว่าเป็น " การวัด " เพราะไม่สามารถบอกปริมาณมากน้อยได้ เป็นแต่เพียงแสดงให้เห็นความแตกต่างของสิ่งของต่าง ๆ เท่านั้น ตัวอย่างการวัดในระดับนี้ เช่น การจำแนกคนตามเพศ เป็นเพศ ชาย-หญิง หรือจำแนกคนตามศาสนาที่นับถือ เป็นคนที่นับถือศาสนาพุทธ คริสต์ อิสลาม เป็นต้น

การกำหนดตัวเลขให้กับสิ่งต่าง ๆ ในมาตรานามบัญญัติจะเป็นเพียงตัวแทนประเภทของที่ถูกวัด หรือเพื่อใช้ในการสื่อความหมาย ให้จำง่ายหรือสะดวกในการนับ โดยที่ตัวเลขดังกล่าวไม่มีความหมายในเชิงปริมาณแต่ประการใด ดังนั้นจึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หรือหารกันได้ นอกจากกระทำได้แต่เพียงการนับจำนวนหรือหาความถี่ของลักษณะที่มีเหมือนกันเท่านั้น เช่น กำหนดตัวเลข 1 แทนเพศชาย และ เลข 2 แทนเพศหญิง

2. มาตราจัดอันดับ (Ordinal Scales) การวัดตามมาตรานี้จะมีระดับการวัดสูงกว่ามาตรานามบัญญัติ ลักษณะการวัดเป็นการจัดอันดับ (rank order) ให้กับคุณสมบัติบางอย่างว่ามากกว่าหรือน้อยกว่าลักษณะอื่น ๆ ซึ่งสามารถจัดอันดับข้อมูลตามตำแหน่งได้จาก

มากที่สุดไปหาน้อยที่สุด หรือจากน้อยที่สุดไปหามากที่สุดได้ เช่น นักเรียนที่สอบได้ที่ 1 , 2 , 3 ,... เป็นต้น ผลการวัดในมาตรานี้ช่วงห่างระหว่างอันดับจะไม่เท่ากัน และไม่ทราบว่าห่างกันเป็นปริมาณเท่าใด เช่น จะบอกว่านักเรียนที่สอบได้ที่ 1 เก่งกว่า นักเรียนที่สอบได้ที่ 2 เท่ากับ นักเรียนที่สอบได้ที่ 3 เก่งกว่านักเรียนที่สอบได้ที่ 4 ไม่ได้ เป็นต้น ดังนั้นข้อมูลที่ปรากฏออกมาเป็นตัวเลขในระดับนี้ไม่สามารถนำมาบวก ลบ คูณ หรือ หารกันได้ จะบอกได้แต่เพียงว่าต่างกันไปในทางไหน หรือทำให้ทราบทิศทางเท่านั้น

3. มาตราอันตรภาค (Interval Scales) มาตราการวัดในมาตรานี้มีลักษณะเหมือนมาตราจัดอันดับทุกอย่าง แต่มีคุณสมบัติเพิ่มเติมมากกว่าตรงที่แต่ละหน่วยของการวัดมีระยะห่างเท่า ๆกัน จึงสามารถเปรียบเทียบกันได้ว่ามากน้อยกว่ากันเท่าใดแต่ไม่สามารถบอกได้ว่าเป็นกี่เท่าของกันและกัน เพราะมาตรานี้ไม่มีศูนย์แท้หรือศูนย์สมบูรณ์ (true zero or absolute zero) เช่น ความร้อนที่มีหน่วยการวัดเป็นฟาเรนไฮท์ และ เซนเซียส ณ. ที่อุณหภูมิ 0 องศาเซนเซียส ไม่ได้หมายความว่า ปราศจากความร้อนเลย ซึ่ง 0 องศาเซนเซียสนี้เท่ากับ 32 องศาฟาเรนไฮท์ หรือผู้ที่สอบได้ศูนย์คะแนน ไม่ได้หมายความว่าไม่มีความรู้ในวิชานั้น เป็นแต่เพียงผู้สอบคนนั้นทำข้อสอบฉบับนั้นไม่ได้เลย เป็นต้น ด้วยเหตุนี้ข้อมูลที่วัดได้ในมาตรานี้จึงไม่สามารถนำมา คูณ หรือ หาร กันได้ จะใช้ได้เฉพาะ บวก หรือ ลบ เท่านั้น

4. มาตราอัตราส่วน (Ratio Scales) เป็นมาตราการวัดในระดับสูงที่สุด กล่าวคือ ครอบคลุมคุณสมบัติทุกอย่างของมาตราอันตรภาค แต่ดีกว่ามาตราอันตรภาคตรงที่มีศูนย์แท้ ข้อมูลในระดับนี้สามารถบวก ลบ คูณ หรือ หารกันได้ มาตราการวัดที่จัดอยู่ในระดับนี้มักจะเป็นการวัดทางฟิสิกส์ เช่น ความยาว ความสูง น้ำหนัก เวลา เป็นต้น

<u>ตอนที่ 2</u> สถิติทดสอบ

<u>สถิติทดสอบเอช ของครัสคัล-วอลลิส</u> (The Kruskal - Wallis' H-Test)

ครัสคัลและวอลลิส เป็นผู้พัฒนาสถิติทดสอบนี้ขึ้นมาในปี ค.ศ.1952 เพื่อศึกษาเปรียบเทียบความแตกต่างอย่างมีนัยสำคัญทางสถิติ ของประชากรซึ่งเป็นที่มาของกลุ่มตัวอย่างสุ่มตั้งแต่สามกลุ่มที่เป็นอิสระต่อกัน สถิติทดสอบ เอช ของครัสคัล-วอลลิส นี้มีพื้นฐานการสร้างเช่นเดียวกับสถิติทดสอบของวิลค็อกซอน (Wilcoxon's Rank-Sum Test) แตกต่างกันที่สถิติทดสอบ เอชนั้นใช้เปรียบเทียบกับกลุ่มตัวอย่างที่มากกว่า 2 กลุ่มขึ้นไป และมีลักษณะ

การวิเคราะห์ในท่านองเดียวกับการวิเคราะห์ความแปรปรวนแบบทางเดียว โดยผลบวกกำลังสองของความแปรปรวนระหว่างกลุ่มนั้น (Mean Square between group) คิดคำนวณโดยใช้ค่าอันดับซึ่งมีค่าเท่ากับ $\sum_{i=1}^{k} n_i(\overline{r}_i - \overline{r})^2$ และค่ากำลังสองของความแปรปรวนทั้งหมด (total mean square) มีค่าเท่ากับ $N(N+1)/12$ ซึ่งสามารถให้นิยามของสถิติทดสอบ เอช ในรูปแบบนี้ได้คือ

$$H = \frac{\sum_{i=1}^{k} n_i(r_i - \overline{r})^2}{1/_{12}\ N(N+1)}$$

หรือในรูปที่นิยมใช้โดยทั่วไปคือ

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R^2}{n_i} - 3(N+1)$$

ครัสคัลและวอลลิส ได้แสดงให้เห็นว่า ถ้ากำหนดให้ r_{ik} แทนอันดับของ x_{ik} และ n แทนขนาดของกลุ่มตัวอย่างแต่ละกลุ่ม

$$N = n_1 + n_2 + n_3 + \ldots\ldots\ldots\ldots\ n_k$$

ดังนั้น $$\overline{R}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} r_{ik} = \frac{R_k}{n_k}$$

และ $$E(\overline{R}_k) = \overline{R}$$

จากนิยามของการกระจายไคสแควร์ สามารถแสดงได้ว่า U ประมาณได้ด้วยการกระจายของไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ k-1 เมื่อ

$$U = Z_1^2 + Z_2^2 + \ldots\ldots\ldots\ldots + Z_k^2 \qquad \ldots\ldots\ldots\ldots(1)$$

$$= \sum_{k=1}^{k} \frac{(R_k - R)^2}{VAR(R_k)}$$

เนื่องจาก $E(r_{ik}) = \frac{N+1}{2}$

$$VAR(r_{ik}) = \frac{N^2 - 1}{12}$$

$$E(R_k) = E\left[\frac{1}{n_k} \sum_{i=1}^{n_k} r_{ik}\right]$$

$$= \frac{1}{n_k} n_k \sum (r_{ik})$$

$$= E(r_{ik})$$

$$= \frac{N+1}{2}$$

$$VAX(R_k) = \frac{VAR(r_{ik})}{n_k} \left[\frac{N - n_k}{N - 1}\right]$$

$$= \frac{1}{n_k} \left[\frac{N^2 - 1}{12}\right] \left[\frac{N - n_k}{N - 1}\right]$$

$$= \frac{(N + 1)\ (N - n_k)}{12n_k}$$

ถ้า n_k มีขนาดใหญ่ ตามทฤษฎี central - limit แล้ว R_k จะมีการกระจายเป็นแบบปกติ

$$Z_k = \frac{R_k - E(R_k)}{\sqrt{VAR(R_k)}}$$

$$Z^2_k = \frac{(R_k - (N + 1)/2)^2 \cdot 12k}{(N + 1)\ (N - n_k)}$$

ซึ่ง Z^2_k จะมีการกระจายประมาณด้วยการกระจายไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ 1

$$\sum_{k=1}^{k} \left[\frac{N - n_k}{N}\right] Z^2_k = \sum_{k=1}^{k} \frac{(N - n_k)}{N} \cdot \left[\frac{R_k - (N + 1)/2}{(N + 1)(N - n_k)}\right]^2 \cdot 12n_k$$

$$= \sum_{k=1}^{k} \frac{12n_k}{N(N + 1)} \left[\bar{R}_k - \frac{(N + 1)}{2}\right]^2$$

$$= \frac{12}{N(N + 1)} \sum_{k=1}^{k} n_k \left[\bar{R}_k - \frac{(N + 1)}{2}\right]^2$$

$$= H$$

ดังนั้น H จะมีการกระจายประมาณได้ด้วยการกระจายของไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ k-1 เมื่อ N มีขนาดใหญ่

ข้อตกลงเบื้องต้นสำหรับ Kruskal-Wallis' H-Test (Conover, 1980)

1. กลุ่มตัวอย่างทั้ง k กลุ่มจะต้องเป็นกลุ่มตัวอย่างสุ่มของแต่ละประชากร
2. กลุ่มตัวอย่างสุ่มภายในกลุ่มและระหว่างกลุ่มเป็นอิสระต่อกัน
3. มาตราการวัดอย่างน้อยที่สุดเป็นมาตราจัดอันดับ
4. ฟังก์ชั่นการกระจายของประชากร k กลุ่มเป็นอันหนึ่งอันเดียวกันโดยประมาณ

การคำนวณค่าสถิติทดสอบ (Test statistics)

1. กรณีที่มีอันดับไม่ซ้ำกัน (untied rank)

นำคะแนนที่ได้จากการสังเกตทั้งหมดมาจัดอันดับแล้วรวมค่าอันดับของแต่ละกลุ่ม ให้ผลรวมของค่าอันดับในแต่ละกลุ่มเป็น R_i แล้วนำไปหาค่าทดสอบ เอช จากสูตรเอช ที่ไม่ใช้ค่าแก้ดังนี้

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

2. กรณีที่มีอันดับที่ซ้ำกัน (ties rank)

นำคะแนนที่ได้จากการสังเกตทั้งหมดมาจัดอันดับ และในอันดับที่ซ้ำกันนั้นให้ใช้ค่าเฉลี่ยของอันดับเป็นตัวแทนอันดับที่ซ้ำ แล้วนับจำนวนความถี่ของการซ้ำในแต่ละอันดับให้เป็น t_s และ รวมค่าอันดับของแต่ละกลุ่มให้ผลรวมของค่าอันดับในแต่ละกลุ่มเป็น R_i แล้วนำไปหาค่าทดสอบ เอช จากสูตร เอช ที่ใช้ค่าแก้ ดังนี้

$$H^* = \frac{\dfrac{12}{N(N+1)} \sum_{i=1}^{k} \dfrac{R_i^2}{n_i} - 3(N+1)}{1 - \dfrac{1}{N^3 - N} \sum_{s=1}^{d} (t_s^3 - t_s)}$$

เมื่อ N คือ จำนวนค่าสังเกตทั้งหมด

n คือ จำนวนค่าสังเกตในแต่ละกลุ่มตัวอย่าง

R คือ ผลรวมของอันดับในแต่ละกลุ่มตัวอย่าง

t คือ จำนวนความถี่ของการซ้ำในแต่ละอันดับ

k คือ จำนวนกลุ่มตัวอย่างทั้งหมด

s คือ จำนวนกลุ่มของอันดับที่มีการซ้ำเกิดขึ้น

การคำนวณค่าวิกฤติสำหรับ Kruskal-Wallis' H-Test

เนื่องจากจำนวนกลุ่มตัวอย่างเท่ากับ 3 (k = 3) และในแต่ละกลุ่มมีจำนวนค่าสังเกตเท่ากับ 5 หรือน้อยกว่า 5 สามารถใช้ค่านัยสำคัญ เอช เทส จากตารางซึ่งสร้างโดยครัสคัลและวอลลิส (Kruskal and Wallis, 1952) ซึ่งมีในหนังสือสถิตินันพาราเมตริกมาตรฐานทุกเล่ม

เมื่อกลุ่มตัวอย่างมีขนาดมากกว่า 15 ค่านัยสำคัญของ เอช เทส สามารถหาจากตารางไคสแควร์ที่ชั้นแห่งความเป็นอิสระเท่ากับ k-1

ลักษณะการกระจายของ Kruskal-Wallis' H-Test

ครัสคัล (Kruskal,1952) ได้พิสูจน์ว่าการกระจายของ เอช เทส ภายใต้สมมติฐานสูญที่เป็นจริงนั้นจะมีลักษณะเป็นไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ k- 1 เอช จะมีค่าเฉลี่ยเท่ากับ k-1 และค่าความแปรปรวนเท่ากับ

$$2(K-1) - \frac{2[3k^2 - 6k + N(2k^2 - 6k+1)]}{5N(N+1)} - \frac{6}{5}\sum_{i=1}^{k}\frac{1}{n_i}$$

เมื่อขนาดของกลุ่มตัวอย่างใหญ่มาก (sample size approaches infinity) ค่าเฉลี่ยและค่าความแปรปรวนของ สถิติทดสอบ เอช จะเท่ากับ k-1 และ 2(k-1) ซึ่งมีค่าเท่ากับค่าเฉลี่ย และ ค่าความแปรปรวนของการกระจายไคสแควร์ ที่ชั้นแห่งความเป็นอิสระ k-1 ซึ่งในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่นี้ครัสคัล และ วอลลิส ได้รับรองว่าการใช้ค่านัยสำคัญของ สถิติทดสอบเอช นี้สามารถประมาณได้จากค่าของไคสแควร์ที่ชั้นแห่งความเป็นอิสระ k-1 ได้ดีที่สุด โดยประมาณว่าค่าสังเกตในแต่ละกลุ่มตัวอย่างมากกว่า 5 ก็สามารถใช้ได้

เมื่อกลุ่มตัวอย่างมีขนาดกลาง ๆ คราสคัลได้แนะนำว่าอาจจะใช้การกระจายโดยประมาณของแกมมา หรือ เบต้า ก็ได้ (Gamma and Beta Distribution) ในกรณีที่ใช้การกระจายโดยประมาณของแกมมา จะมีค่าเฉลี่ย (E) คือ $E = k-1$

และค่าความแปรปรวน (V) คือ

$$V = 2(K-1) - \frac{2[3k^2 - 6k + N(2k^2 - 6k + 1]}{5N(N+1)} - \frac{6}{5}\sum_{i=1}^{k}\frac{1}{n_i}$$

และแปลงเป็นค่าไคสแควร์คือ $X^2 = 2HE/V$ ที่มีค่าชั้นแห่งความเป็นอิสระเท่ากับ $2E^2/V$ กรณีที่ใช้การกระจายโดยประมาณของเบต้าใช้ค่าเฉลี่ย (E) และค่าความแปรปรวน (V) เท่ากับค่าการกระจายโดยประมาณของแกมมาและแปลงเป็นการกระจายแบบเอฟ ดังนี้

$$F = \frac{H(M - E)}{E(M - H)}$$ โดย

$$M = \frac{N^3 - \sum_{i=1}^{k} n_i^3}{N(N+1)}$$ และจำนวนชั้นแห่งความเป็นอิสระเท่ากับ

$$f_1 = \frac{E(M - E) - V}{1/2\, MV} \quad \text{และ} \quad f_2 = \frac{M - E}{E} \cdot f_1$$

แต่ค่าชั้นแห่งความเป็นอิสระที่ได้ไม่เป็นจำนวนเต็ม ทำให้ไม่สามารถใช้ค่าเอฟได้จึงแปลงเป็นการกระจายแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 คือ

$$k_p = \frac{(1 - 2/9f_2)F + 2/9f_1 - 1}{\sqrt{2f^2/9f_2 + 2/9f_1}}$$

และใช้ค่าความน่าจะเป็นจากตารางการกระจายแบบปกติ

การกระจายที่แท้จริงของเอชภายใต้สมมติฐานสูญ และการกระจายโดยประมาณด้วยการกระจายไคสแควร์, แกมมาและเบต้านั้นได้แสดงการเปรียบเทียบการกระจายทั้ง 4 ชนิดนี้ในแผนภาพที่ 2 และ 3 (Kruskal and Wallis, 1952) โดย X^2 แทนการกระจายของไคสแควร์ Γ แทนการกระจายของแกมมา และ β แทนการกระจายของเบต้า

<u>การคำนวณค่าสถิติทดสอบเอชเมื่อข้อมูลเป็นมาตราส่วนประมาณค่า</u>

การคำนวณค่าสถิติทดสอบเอช เมื่อข้อมูลที่รวบรวมได้เป็นลักษณะมาตราส่วนประมาณค่า ทำได้โดยจัดข้อมูลในรูปตารางการณ์จร ดังนี้คือ

Population Category	1	2	3	...	k	ROW Totals	R_i = Average Rank
1	O_{11}	O_{12}	O_{13}	...	O_{1k}	t_1	$(t_1+1)/2$
2	O_{21}	O_{22}	O_{23}	...	O_{2k}	t_2	$t_1+(t_2+1)/2$
3	O_{31}	O_{32}	O_{33}	...	O_{3k}	t_3	$t_1+t_2+(t_3+1)/2$
...	...	...	...	...	...		
C	O_{c1}	O_{c2}	O_{c3}	...	O_{ck}	t_c	$\sum_{i=1}^{c-1} t_i+(t_c+1)/2$
Column Totals	n_1	n_2	n_3	...	n_k	N = Grand Total	

จากนั้นคำนวณหาค่าเฉลี่ยของอันดับเพื่อเป็นตัวแทนอันดับที่ซ้ำ แล้วจึงหาผลรวมของอันดับในแต่ละกลุ่มและ นำผลรวมของอันดับในแต่ละกลุ่มไปแทนค่าในสูตร H^* และนำค่าที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤตจากตารางของ X^2 ที่ชั้นแห่งความเป็นอิสระเท่ากับ k-1

เราสามารถจะตรวจสอบผลรวมของอันดับในแต่ละสดมภ์ได้จาก $R_j = \sum_{i=1}^{c} O_{ij} R_i$

การทดสอบการแจกแจงด้วยไคสแควร์

ประวัติและพัฒนาการของไคสแควร์

Chi อ่านว่า ไค เป็นคำที่มาจากภาษากรีก ใช้สัญลักษณ์ X แทน แต่เดิมไม่ได้เรียกว่าการแจกแจงแบบไคสแควร์ (Cochran, 1952) เนื่องจากปีคริสตศักราช 1936 ได้มีชาวเยอรมันชื่อ เฮลเมอร์ท (Robert Ericdrich Helmert) ค้นพบการแจกแจงที่เป็นโค้งปกติ และเรียกว่าการแจกแจงแบบ เฮลเมอร์ท (Helmert's Distribution) การกระจายซึ่งเฮลเมอร์ทค้นพบได้ถูกละเลยเป็นเวลา 20 ปีเศษ

ต่อมาในปีคริสตศักราช 1900 เปียร์สัน (Karl Pearson) ได้ค้นพบใหม่และผลงานชิ้นนี้นับว่าเป็นรากฐานสำหรับสถิติแผนใหม่ โดยเริ่มจากการพิสูจน์ว่าถ้ากลุ่มของตัวแปรที่สัมพันธ์กัน Z; (Correlate Variates) จำนวน n ตัว มีมัชฌิมเลขคณิตเท่ากับศูนย์ มีการแจกแจงความน่าจะเป็นดังนี้

$$dF = (2\pi)^{-1/2^{n}} \quad \exp^{(Q/2)} \quad dz1\ dz2\ dz3 \ldots dz_n$$

(charles C.Peter and Walter Van Vourhis, 1940)

หลังจากนั้นเปียร์สัน (Karl Pearson) ได้เสนอการทดสอบภาวะสารูปสนิทดี ที่ข้อมูลแยกออกจากกันโดยเริ่มจากการทราบค่าความถี่ที่คาดหวัง (m_i) ล่วงหน้าและ x_i ใด ๆ ที่ได้จากการสังเกตจะต้องมีการแจกแจงแบบมัลติโนเมียล (Multinomial Distribution) เปียร์สันยอมรับว่า x_i นั้นอาจมีการแจกแจงเป็นโค้งปกติ ซึ่งจุดนี้เปียร์สันยอมรับว่าในกรณีที่ค่าความถี่ที่คาดหวังมีขนาดใหญ่พอ ค่า และผลงานในตอนแรกของเขานี้ ทำให้ทราบว่าถ้าตัวอย่างมีขนาดใหญ่ค่า X^2 จะมีชั้นแห่งความเป็นอิสระเท่ากับ k-1

ต่อมาฟิชเชอร์ (R.A Fisher) ได้หลีกเลี่ยงความยุ่งยากในงานของเปียร์สัน โดยชี้ให้เห็นว่าถ้ายอมรับว่าสิ่งที่สังเกตได้ x_i มีการแจกแจงแบบพัวซอง (Poisson Distribution)

$$\prod_{i=1}^{k} \frac{e^{-m} m_i^{x_i}}{x_i!} = e^{-n} \prod_{i=1}^{k} \frac{m_i^{x_i}}{x_i!} , \quad m = n$$

มีค่า m_i เป็นค่าความถี่ที่คาดหวังหรือมัชฌิมและขนาดของกลุ่มตัวอย่าง n ใหญ่มากแล้ว

$Y_i = \dfrac{x_i - m_i}{\sqrt{m_i}}$ จะมีการแจกแจงโค้งปกติ เมื่อมัชฌิมเลขคณิตคือ m_i และส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\sqrt{m}$ ดังนั้นการแจกแจงของ X^2 คือ $Y_1^2 + Y_2^2 \dots Y_k^2$ เมื่อ Y_i มีการแจกแจงอย่างอิสระ และ

$$\sum_{i=1}^{k} Y_i \sqrt{m_i} = \sum_{i=1}^{k} (x_i - m_i) = 0$$

<u>การแจกแจงของไคสแควร์</u>

โดยปกติแล้ว การใช้ไคสแควร์ ในการทดสอบจะสัมพันธ์กับการเปรียบเทียบการแจกแจงข้อมูลที่ต่อเนื่อง แต่ในทางปฏิบัติมักจะใช้กับการแจกแจงของข้อมูลที่ไม่ต่อเนื่อง เช่น ข้อมูลที่เป็นความถี่ ทำให้โค้งขาดตอนซึ่งก็ถือว่ามีการแจกแจงแบบไคสแควร์ ลักษณะการแจกแจงของไคสแควร์นั้นขึ้นอยู่กับจำนวนชั้นแห่งความเป็นอิสระ ถ้าจำนวนชั้นแห่งความเป็นอิสระมาก ก็จะมีการแจกแจงที่เบ้น้อย โดยใช้สูตรในการคำนวณดังนี้

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad \text{.............(1)}$$

O_i = ความถี่ที่ได้จากการสังเกต (Observation Frequency)

E_i = ความถี่ที่คาดหวัง (Expected Frequency)

แต่สมการส่วนโค้งการกระจายของไคสแควร์ที่แท้จริงเป็นดังนี้

$$Y = \frac{1}{(n/2-1)\Gamma 2^{n/2}} (X^2)^{(n/2-1)} e^{-\frac{1}{2}X^2} \qquad \text{...........(2)}$$

ทั้งสูตร (1) และ (2) นั้นเมื่อหาค่าความน่าจะเป็นแล้วจะมีค่าใกล้เคียงกันมาก เปียร์สัน (Karl Pearson) ได้แสดงให้เห็นว่าค่าที่ได้จากทั้ง 2 สูตรนั้นมีสหสัมพันธ์กันอยู่ระหว่าง .93 - .99 ซึ่งนับว่าสูงมาก (Peter and Voorhis, 1940)

ดังนั้นจึงได้นำไคสแควร์ไปใช้กันอย่างกว้างขวาง ไม่ว่าข้อมูลเหล่านั้นจะเป็นความถี่ที่ต่อเนื่องหรือไม่ต่อเนื่องก็ได้ แต่หลักใหญ่คือ การเปรียบเทียบความแตกต่างของความถี่ที่คาดหวัง กับความถี่ที่ได้จากการสังเกต ซึ่งในที่นี้ความถี่ที่คาดหวังจะเป็นตัวพารามิเตอร์ และความถี่ที่สังเกตได้จะเป็นค่าประมาณจากพารามิเตอร์ ถ้าทั้งสองค่านี้ใกล้เคียงกันการทดสอบจะไม่มีนัยสำคัญ

เหตุผลสำคัญที่ใช้สูตรไคสแควร์ในรูป $\sum_{i=1}^{k} (O_i - E_i)^2 / Ei$ ก็เพราะเป็นผลบวกของค่าประมาณ k ตัว ที่ได้มาจากโค้งปกติซึ่งจะทำให้ค่า $(O_i - E_i)/E_i$ มีการแจกแจงเป็นโค้งปกติมีมัชฌิมเท่ากับ 0 และส่วนเบี่ยงแบนมาตรฐานเท่ากับ 1 เหมือนกับลักษณะการแจกแจงไคสแควร์ที่แท้จริง ดังนั้นจึงสามารถใช้แทนกันได้

<u>การทดสอบการแจกแจงด้วยไคสแควร์</u> (Chi-Square test of Homogeneity of Distributions)

การทดสอบลักษณะการแจกแจงของประชากรตั้งแต่ 2 กลุ่มขึ้นไป โดยที่ข้อมูลในลักษณะการจัดกลุ่มหรือจัดประเภทได้มากกว่า 2 ประเภท (Multinomial) ว่าลักษณะการแจกแจงเป็นเช่นเดียวกันทุกกลุ่มหรือไม่สามารถใช้ การทดสอบการแจกแจงด้วยไคสแควร์ ซึ่งคำนวณจากสูตรดังต่อไปนี้

$$X^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad df = (c-1)(r-1)$$

เมื่อ O คือ ค่าความถี่ที่สังเกตได้

E คือ ค่าความถี่ที่คาดหวัง

C คือ จำนวนกลุ่มตัวอย่าง

R คือ จำนวนประเภทของข้อมูล

ข้อจำกัดและข้อตกลงเบื้องต้นนี้คือ

1. กลุ่มตัวอย่างระหว่างประชากรเป็นอิสระต่อกัน (Independent between Sample)

2. กลุ่มตัวอย่างภายในประชากรเป็นอิสระต่อกัน (Independent with in sample)

3. ข้อมูลในลักษณะจัดกลุ่มหรือจัดประเภทได้มากกว่า 2 ประเภท (Multinomial)

4. ค่าความถี่ที่คาดหวังทุกช่องจะต้องมีค่ามากกว่า 5 (Expected value greater than 5)

วิธีการดำเนินการทดสอบ จะต้องหาค่าความถี่ที่คาดหวังก่อนแล้วคำนวณ

$$X^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

แล้วเปรียบเทียบกับค่าวิกฤตจากตารางของ X^2 ตาม α ที่ระบุไว้ตาม degree of freedom ซึ่งจะได้จากการคำนวณ df = (C-1)(r-1)

เพื่อความสะดวกในการเข้าใจวิธีการหาค่าความถี่ที่คาดหวังจึงขอแสดงด้วยตารางดังต่อไปนี้

MULTINOMIAL	1	2	3	.	.	.	j	รวม
1	O_{11}	O_{12}	O_{13}	-	-	-	O_{1j}	$O_{1.}$
2	O_{21}	O_{22}	O_{23}	-	-	-	O_{2j}	$O_{2.}$
3	O_{31}	O_{32}	O_{33}	-	-	-	O_{3j}	$O_{3.}$
.								
.								
i	O_{i1}	O_{i2}	O_{i3}	-	-	-	O_{ij}	$O_{i.}$
รวม	$O_{.1}$	$O_{.2}$	$O_{.3}$	-	-	-	$O_{.j}$	N

คำนวณค่าความถี่ที่คาดหวังดังต่อไปนี้

$$E_{11} = \frac{O_{.1} \times O_{1.}}{N}$$

$$E_{12} = \frac{O_{.2} \times O_{1.}}{N}$$

.

.

.

$$E_{1j} = \frac{O_{.j} \times O_{1.}}{N}$$

E_{ij} แทนที่ความถี่ที่คาดหวังแถวที่ i สดมภ์ที่ j

$O_{.j}$ แทนผลรวมของค่าความถี่ที่สังเกตได้ทางด้านแถวของสดมภ์ที่ j

$O_{i.}$ แทนผลรวมของค่าความถี่ที่สังเกตได้ทางด้านสดมภ์ของแถวที่ j

N แทนจำนวนความถี่ทั้งหมด

การคำนวณสถิติทดสอบไคสแควร์เมื่อข้อมูลเป็นมาตราส่วนประมาณค่า

การคำนวณค่าสถิติทดสอบไคสแควร์ เมื่อข้อมูลที่รวบรวมได้เป็นลักษณะมาตราส่วนประมาณค่า ทำได้โดย จัดข้อมูลในรูปตารางการจรดังนี้คือ

Population Category	1	2	3	...	k	ROW Totals
1	O_{11}	O_{12}	O_{13}	...	O_{1k}	t_1
2	O_{21}	O_{22}	O_{23}	...	O_{2k}	t_2
3	O_{31}	O_{32}	O_{33}	...	O_{3k}	t_3
...	...	...	...	...	...	
C	O_{c1}	O_{c2}	O_{c3}	...	O_{ck}	t_c
Column Totals	n_1	n_2	n_3	...	n_k	N = Grand Total

จากนั้นคำนวณหาค่าความถี่ที่คาดหวังในแต่ละเซลก่อน แล้วคำนวณหาค่าไคสแควร์ นำค่าไคสแควร์ที่คำนวณได้ เปรียบเทียบกับค่าวิกฤตจากตารางของ X^2 ตาม α ที่ระบุไว้ ตาม degree of freedom

ระดับความมีนัยสำคัญความคลาดเคลื่อนประเภทที่ 1และประเภทที่ 2และอำนาจของสถิติทดสอบ

ในการทดสอบสมมติฐานนั้นผู้ทดสอบจะมีโอกาสตัดสินใจอยู่ 2 ลักษณะคือ ปฏิเสธสมมติฐานสูญกับยอมรับ สมมติฐานสูญ ซึ่งการตัดสินอย่างใดอย่างหนึ่งนั้น ต่างก็ต้องเสี่ยงต่อความคลาดเคลื่อนกับสภาพความเป็นจริงอยู่ ซึ่งงานทางสถิติได้ให้ความหมายของความคลาดเคลื่อน ตลอดจนศัพท์ที่เกี่ยวข้องกันกับความคลาดเคลื่อน (Glass, 1970)ไว้ดังนี้คือ

ระดับความมีนัยสำคัญ คือ โอกาสที่ยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 แทนด้วยสัญลักษณ์ α

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) คือ ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐานสูญที่ถูก โอกาสที่จะทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 กำหนดด้วย α

ความคลาดเคลื่อนประเภทที่ 2 (Type I error) คือ ความคลาดเคลื่อนที่เกิดจากการยอมรับสมมติฐานสูญที่ผิด โอกาสที่จะทำให้เกิดความคลาดเคลื่อนประเภทที่ 2 กำหนดด้วย β

อำนาจการทดสอบ (Power of Test) คือ ความน่าจะเป็นที่จะปฏิเสธสมมติฐานสูญ เมื่อสมมติฐานสูญนั้นผิด ซึ่งมีค่าเท่ากับ $(1-\beta)$

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐานสูญเมื่อสมมติฐานสูญเป็นจริง ความคลาดเคลื่อนประเภทนี้จะเกิดขึ้นในกรณีที่ผู้ทดสอบสมมติฐานปฏิเสธสมมติฐานสูญเท่านั้น โอกาสที่จะยอมให้เกิดความคลาดเคลื่อนประเภทนี้ ผู้ใช้สถิติจะระบุไว้ล่วงหน้าที่จุดเสี่ยงในโอกาสประมาณเท่าไรซึ่งรู้จักกันในทั่วไปว่าคือระดับความมีนัยสำคัญ มักจะใช้สัญลักษณ์ α

ความคลาดเคลื่อนประเภทที่ 2 (Type II error) เป็นความคลาดเคลื่อนที่เกิดจากการยอมรับสมมติฐานสูญที่ผิด ความคลาดเคลื่อนประเภทนี้จะเกิดขึ้นในกรณีที่ผู้ทดสอบยอมรับสมมติฐานสูญ ทั้ง ๆ ที่ความจริงแล้วสมมติฐานสูญนั้นผิด ผู้ใช้สถิติทดสอบอาจคำนวณค่าความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 2 ได้เช่นเดียวกัน เมื่อเขาสามารถประมาณค่าความแตกต่างระหว่างพารามิเตอร์ที่ทดสอบได้โอกาสที่อาจจะเกิดความคลาดเคลื่อนประเภทที่ 2 จะเท่ากับ β

ค่า 1-β จะเป็นค่าโอกาสของการทดสอบที่ตัดสินได้ถูกต้องซึ่งทางด้านสถิติจะให้ความหมายค่านี้คือ อำนาจการทดสอบ (Power of test) สถิติทดสอบที่เหมาะสมควรจะมีอำนาจการทดสอบเข้าใกล้ 1.00 หรือ 100%

<u>ความสัมพันธ์ระหว่าง α กับ β ในกรณีที่ Sample Size คงที่</u>

โอกาสที่จะเกิดความคลาดเคลื่อนทั้ง 2 ประเภท นี้มีส่วนเกี่ยวข้องสัมพันธ์กัน คือ ถ้าพยายามลดขนาดของ α ก็จะเป็นการเพิ่มขนาดของ β ถ้าเพิ่มขนาดของ α ก็จะเป็นการลดขนาดของ β ดังนั้นในการทดสอบสมมติฐาน ผู้ทดสอบจะต้องคำนึงด้วยว่า ความคลาดเคลื่อนชนิดใด เป็นความคลาดเคลื่อนที่เขาต้องการจะจำกัดให้เท่ากับหรือน้อยกว่าที่ต้องการซึ่งปกติแล้ว ผู้ใช้สถิติจะนิยมกำหนดโอกาสความคลาดเคลื่อนประเภทที่ 1 เป็นเบื้องต้นและทำกันเป็นประจำ ซึ่งมักพบเห็นได้ในรายงานการวิจัยทั่ว ๆ ไปที่ระบุค่า α

การเพิ่มขนาดของกลุ่มตัวอย่าง ก็มีส่วนช่วยลดโอกาสที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 และ 2 ลงได้

<u>ตอนที่ 3</u> งานวิจัยที่เกี่ยวข้อง

Robeson (1989) ได้ศึกษาเปรียบเทียบการวิเคราะห์ข้อมูลเมื่อข้อมูลเป็นลักษณะ polychotomous data ด้วยการวิเคราะห์ความแปรปรวน (ANOVA) และสถิติทดสอบเอชของ ครัสคัล-วอลลิส โดยใช้วิธีมอนติคาร์โล ซิมูเลชั่น เมื่อกลุ่มตัวอย่างมีขนาด 5,10 และ 15 มีจำนวนกลุ่มตัวอย่าง 2 และ 3 กลุ่ม ข้อมูลมีลักษณะการแจกแจงแบบปกติ และไม่ปกติ โดยมีค่าเฉลี่ยและความแปรปรวนเท่ากัน ผลปรากฏว่า เมื่อลักษณะการแจกแจงของข้อมูลเหมือนกัน และความแปรปรวนเท่ากัน สถิติทดสอบเอชของ ครัสคัล-วอลลิส ใช้ได้ผลดีกว่าการวิเคราะห์ความแปรปรวน (ANOVA) โดยเฉพาะอย่างยิ่งเมื่อจำนวนกลุ่มตัวอย่างเป็น 2 หรือ 3 กลุ่ม และขนาดของกลุ่มตัวอย่างเป็น 5 หรือ 10

Blair and Higgins (1980) ได้ใช้เทคนิคมอนติคาร์โลทำการศึกษา เพื่อเปรียบเทียบอำนาจการทดสอบของสถิติทดสอบวิลคอกซอน กับสถิติทดสอบที ภายใต้ลักษณะการแจกแจงของประชากรไม่เป็นแบบปกติ ผลการศึกษาสรุปได้ว่า โดยทั่วไปสถิติทดสอบของวิลคอกซอน มีอำนาจการทดสอบสูงกว่าสถิติทดสอบทีมาก

ไปรมา พจนพิมล (2525)ได้ใช้เทคนิคมอนติคาร์โล ทำการศึกษาเพื่อเปรียบเทียบอำนาจการทดสอบของสถิติทดสอบที (t-test) สถิติทดสอบของวิลคอกซอน (Wilcoxon test), สถิติทดสอบของเทอรี่-โฮฟฟ์ดิง (Terry-Hoeffding Normal-Scores Test)

และสถิติทดสอบของเดอ แวเดน (Van der Waerden Normal Scores Test), ได้ผลสรุปว่า ถ้าข้อมูลที่ใช้ในการวิเคราะห์อยู่ในรูปของอันดับ (rank) สถิติทดสอบของเทอรี่-โฮฟฟ์ดิง และสถิติทดสอบของแวน เดอ แวร์เดน มีอำนาจการทดสอบสูงกว่าสถิติทดสอบของวินคอกซอน

Boneau (1962) ได้ใช้เทคนิคมอนติคาร์โลศึกษาเปรียบเทียบอำนาจการทดสอบของสถิติทดสอบวิบคอกซอน เมื่อกลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็น normal, rectangular และ exponential ได้สรุปว่าโดยทั่วไปแล้วสถิติทดสอบที มีอำนาจการทดสอบสูงกว่าสถิติทดสอบ Mann-Witney U test (Wilcoxon test) แต่ไม่มากนัก

Williams (1950) ได้ศึกษาในกรณีกลุ่มตัวอย่างน้อยมาก ๆ พบว่า ไคสแควร์ เทส ไม่สามารถจะนำมาใช้ได้ เนื่องจากอาจจะทำให้เกิดความคลาดเคลื่อนได้พร้อมกับจะต้องมีวิธีการในการทดสอบมากขึ้น ดังนั้นไคสแควร์ เทส จึงควรใช้กับขนาดของกลุ่มตัวอย่างที่ใหญ่ขึ้น ซึ่งกรณีกลุ่มตัวอย่างน้อย ๆ ถ้าใช้คอลมอโกรอฟ สเมอร์นอฟ เทส จะทำให้อำนาจของการทดสอบสูงกว่าไคสแควร์ เทส

Slakter (1965:854-858) ได้ศึกษาเปรียบเทียบระหว่างไคสแควร์ เทส และคอลมอโกรอฟ สเมอร์นอฟ เทส ซึ่งเป็น Goodness-of-fit test ทั้งคู่ภายใต้เงื่อนไขดังนี้

1. จำนวนข้อมูลที่สังเกตได้ (Observations) N จำนวนต้องเป็นอิสระซึ่งกันและกัน ในขณะที่ k กลุ่มต้องแยกจากกันโดยเด็ดขาด แต่โอกาสภายใต้สมมติฐานในการทดสอบควรจะเท่ากัน

2. ทั้ง N และ k ต้องมีขนาดเล็กไม่ควรมากกว่า 50

จากกลุ่มตัวอย่างที่ถูกสุ่มมาทำการทดลองปรากฏว่า คอลมอโกรอฟ สเมอร์นอฟ เทส มีอำนาจของการทดสอบดีกว่าไคสแควร์ เทส เมื่อขนาดของกลุ่มตัวอย่างเล็ก ๆ ขณะที่ไคสแควร์ เทส ต้องมีอำนาจของข้อมูลที่สังเกตได้มากพอที่จะคาดประมาณค่าได้ในการทดสอบทางสถิติ

Dixon (1954) ได้ทำการศึกษาเปรียบเทียบเค-เอส เทส (K-S test) กับแรงค์ซัม เทส (rank sum test), มีเดียน เทส (median test), รันส์ เทส (runs test) และที เทส (t-test) แบบสองทาง ในกรณีที่มีลักษณะเป็นแบบปกติ (normal shift alternatives) ผลปรากฏว่าดิสตรีบิชั่นฟรี (distribution-Free) ทั้ง 4 วิธี มีอำนาจในการทดสอบสูงกว่าการทดสอบทีเมื่อกลุ่มตัวอย่างมีขนาดเล็ก ซึ่งในการศึกษาเปรียบเทียบนี้ใช้ขนาดของกลุ่มตัวอย่างเท่ากับ 5 มีลักษณะเป็นแบบปกติ ที่ระดับนัยสำคัญ 0.025 สรุปได้ว่าการทดสอบโดยใช้ลำดับ (rank test) มีอำนาจของการทดสอบดีกว่า เค-เอส เทส ส่วนเค-เอส เทส จะมีอำนาจการทดสอบดีกว่ามีเดียน เทส