บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันงานวิจัยในสาขาต่าง ๆ จำเป็นต้องอาศัยความรู้ทางด้านระเบียบวิธีการทางสถิติเข้ามาช่วยในการสรุปผลและตัดสินใจเป็นอย่างมาก โดยเฉพาะอย่างยิ่งการอ้างอิงด้านการทดสอบสมมติฐาน ทั้งนี้เพราะงานวิจัยส่วนใหญ่ไม่สามารถที่จะศึกษาจากประชากรโดยตรงทั้งหมดได้เนื่องจากขนาดของประชากรที่สนใจมักมีขนาดใหญ่มากในแต่ละโครงการวิจัย ดังนั้นในทางปฏิบัติผู้วิจัยมักจะสุ่มตัวอย่างจากประชากรที่สนใจมาเพียงบางส่วน แล้วจึงศึกษาข้อมูลจากตัวอย่าง เพื่ออนุมานคุณสมบัติของประชากรจากคุณสมบัติของตัวอย่างที่ได้ ด้วยเหตุนี้ความถูกต้องและเชื่อถือได้ในการอ้างอิงด้านการทดสอบสมมติฐาน จึงขึ้นอยู่กับการตัดสินใจเลือกใช้สถิติทดสอบ (test statistic) ที่เหมาะสม ซึ่งสถิติทดสอบที่ใช้ในปัจจุบันแบ่งได้เป็น 2 ประเภทใหญ่ ๆ คือ สถิติทดสอบพาราเมตริก (parametric tests) และสถิติทดสอบนันพาราเมตริก (nonparametric tests)

สำหรับสถิติทดสอบพาราเมตริกที่คุ้นเคยกันมาก เช่น สถิติทดสอบที (t-test) หรือ สถิติทดสอบเอฟ (F-test) ในการวิเคราะห์ความแปรปรวน ซึ่งส่วนใหญ่มักจะนำมาใช้ทดสอบสมมติฐานที่เกี่ยวกับค่าเฉลี่ย (μ) ของประชากรกลุ่มเดียว สองกลุ่ม และ มากกว่าสองกลุ่ม โดยที่ลักษณะของข้อมูลที่จะนำมาวิเคราะห์นั้น จะต้องจัดอยู่ในมาตราวัดอันตรภาคเป็นอย่างต่ำจึงจะสามารถหาค่าเฉลี่ย ($\overline{X}$) ของกลุ่มตัวอย่างจากข้อมูลเหล่านั้นได้ แต่สำหรับสถิติทดสอบนันพาราเมตริก เช่น สถิติทดสอบไคสแควร์ (χ^2-test) วิลคือกซอน เทส (Wilcoxon test) เอช เทส ของคราสคัลวอลลิส (The Kruskal Wallis'H-test) หรือ ฟรีดแมน เทส (Friedman test) สามารถนำมาใช้ทดสอบสมมติฐาน เมื่อข้อมูลจัดอยู่ในมาตรานามบัญญัติ (Nominal scale) และมาตราจัดอันดับ (Ordinal scale) ได้

ซึ่งเมื่อพิจารณาด้านระดับการวัดของข้อมูลแล้ว พบว่า ข้อมูลที่วัดได้จากงานวิจัยทางด้านสังคมศาสตร์ในปัจจุบัน ส่วนใหญ่จะเป็นข้อมูลที่จัดอยู่ในมาตรานามบัญญัติ หรือมาตราจัดอันดับเท่านั้น (พิชิต พิทักษ์เทพสมบัติ, 2526)

ส่วนทางด้านการเลือกใช้สถิติทดสอบประเภทใดในการทดสอบสมมติฐานนั้น Siegel (1956) กล่าวว่า เกณฑ์ที่สำคัญประการหนึ่งซึ่งผู้วิจัยควรนำมาพิจารณาในการตัดสินใจเลือกใช้สถิติทดสอบก็คือ ระดับการวัดของข้อมูล โดยเฉพาะอย่างยิ่งข้อมูลในมาตราจัดอันดับนั้นมีผู้ที่สนใจให้ความสำคัญพร้อมกับการให้ข้อสังเกตประกอบคำแนะนำไว้ดังนี้

Siegel (1956) เน้นว่าในงานวิจัยทางด้านพฤติกรรมศาสตร์ ข้อมูลส่วนใหญ่จะวัดได้ถึงมาตราจัดอันดับเท่านั้น ด้วยเหตุนี้ค่าสถิติที่เหมาะสม ซึ่งจะนำไปใช้แทนค่ากลางของข้อมูลก็คือ ค่ามัธยฐาน (median) และสถิติทดสอบที่เหมาะสมซึ่งจะนำไปใช้ในการทดสอบสมมติฐานก็คือ สถิติทดสอบประเภทนันพาราเมตริก สำหรับเหตุผลที่ไม่ควรนำสถิติทดสอบประเภทพาราเมตริก ไปใช้กับข้อมูลที่อยู่ในมาตราจัดอันดับ ก็เพราะว่าการทดสอบสมมติฐานของสถิติทดสอบประเภทพาราเมตริก จะได้ผลสรุปที่ถูกต้องและมีความตรง (validity) ก็ต่อเมื่อข้อมูลที่นำมาวิเคราะห์อย่างน้อยที่สุดต้องอยู่ในมาตราอันตรภาค และเป็นไปตามข้อตกลงเบื้องต้นของสถิติทดสอบนั้นๆ ซึ่งในทางตรงกันข้าม สถิติทดสอบประเภทนันพาราเมตริกสามารถนำไปใช้ได้กับข้อมูลที่อยู่ในมาตราจัดอันดับ และจะไม่เข้มงวดเกี่ยวกับข้อตกลงเบื้องต้นอีกด้วย

Senders (1958) ได้ยกตัวอย่างข้อมูลที่จัดอยู่ในมาตราจัดอันดับ เช่น สเกลวัดทัศนคติ ระดับผลการเรียน และมาตราส่วนประมาณค่าแบบต่าง ๆ ซึ่งความเข้าใจผิดที่พบเห็นบ่อยที่สุดคือ การหาค่าเฉลี่ยเลขคณิต (arithmetic average) กับข้อมูลที่อยู่ในมาตราจัดอันดับ ซึ่งตัวอย่างที่เห็นได้อย่างชัดเจนก็คือ การหาค่าเกรดเฉลี่ย (grade point - averages) ทั้งนี้เพราะ ตัวเลขที่ใช้แทนระดับผลการเรียน (ดีมาก-4 ดี-3 ปานกลาง-2

ผ่าน-1 ตก-0) จัดว่าอยู่ในมาตราจัดอันดับเท่านั้น อีกทั้งยังไม่อาจรับประกันได้ว่าความแตกต่างระหว่างเกรด 4 กับเกรด 3 จะมีค่าเท่ากับ ความแตกต่างระหว่างเกรด 3 กับเกรด 2 ซึ่งผลที่ตามมาก็คือ อาจทำให้การแปลความหมายจากผลการวิเคราะห์ข้อมูลบิดเบือนไปจากความเป็นจริง นอกจากนี้ Senders ยังได้แนะนำหลักทั่วๆ ไปว่า สถิติทดสอบทางด้านพาราเมตริก เช่น ที เทส (t-test) หรือ เอฟ เทส (F-test) ในการวิเคราะห์ความแปรปรวน ไม่เหมาะสมที่จะนำไปใช้กับข้อมูลที่จัดอยู่ในมาตราจัดอันดับ ในทางตรงกันข้าม สถิติทดสอบดังกล่าวข้างต้นนั้น ควรจะนำไปใช้กับข้อมูลที่จัดอยู่ในมาตราอันตรภาค หรือมาตราอัตราส่วนเท่านั้น สำหรับข้อมูลที่จัดอยู่ในมาตราจัดอันดับ สถิติทดสอบทางด้านนันพาราเมตริก เช่น Mann-Whitnet U หรือ Kruskal Wallis จะนำไปใช้ทดสอบได้เหมาะสมกว่า

Adams, Fagot และ Robinson (1965, อ้างถึงใน Knapp, 1990) เสนอแนะว่า ในงานวิจัยทางด้านจิตวิทยาถ้าข้อมูลที่เก็บรวบรวมได้จัดอยู่ในมาตราจัดอันดับแล้ว ผู้วิจัยก็ตัดสินใจเลือกใช้สถิติทดสอบทางด้านนันพารา เมตริกในการทดสอบสมมติฐานได้ ทั้งนี้มิใช่เพราะลักษณะของข้อมูลเพียงอย่างเดียว แต่เหตุผลอีกประการหนึ่งก็คือ สถิติทดสอบทางด้านนันพาราเมตริกนี้จะใช้ทดสอบสมมติฐานได้อย่างเหมาะสม และผลการทดสอบจะมีความหมายสำหรับตัวแปรมาตราจัดอันดับที่วัดได้จากงานวิจัยด้านนี้อีกด้วย

Wiersma (1969) กล่าวว่าสเกลวัดทัศนคติเช่น ลิเคอร์ทสเกล (Likert scale) ซึ่งประกอบด้วยข้อคำถามที่แต่ละข้อ ได้กำหนดคำตอบไว้ให้ผู้ตอบพิจารณาว่า คำตอบนั้น ๆ ควรมีน้ำหนักอยู่ในเกณฑ์ขนาดใด และใช้วิธีการกำหนดคะแนนแทนข้อความ เช่น มากที่สุด มีค่าเท่ากับ 5 คะแนน มาก มีค่าเท่ากับ 4 คะแนน ปานกลาง มีค่าเท่ากับ 3 คะแนน น้อยมีค่าเท่ากับ 2 คะแนน น้อยที่สุด มีค่าเท่ากับ 1 คะแนน โดยที่แต่ละช่วงของคะแนนจะถือเอาว่ามีค่าเท่า ๆ กัน แต่อย่างไรก็ตาม คะแนนแต่ละช่วงที่ต่างกันนี้ แสดงให้เห็นเพียงแค่ว่าผู้ตอบให้น้ำหนักของคำตอบอยู่ในระดับ หรือเกณฑ์ขนาดใดเท่านั้น ซึ่งคะแนนเหล่านี้ไม่สามารถบอกได้อย่างแน่ชัดว่า ระหว่างสองช่วง หรือสองจำนวนที่จัดเรียงลำดับกัน

อยู่นั้นมีความแตกต่างกันอยู่เท่าไร จะทราบได้แต่เพียงว่า แตกต่างกันในลักษณะมากกว่าหรือน้อยกว่ากันเท่านั้น นอกจากนี้คะแนนที่แตกต่างกัน 1 หน่วย ไม่ได้แทนปริมาณที่เท่ากันเหมือนคะแนนในมาตราอันตรภาค หรือมาตราอัตราส่วน และถึงแม้ว่าจะนำคะแนนของผู้ตอบแต่ละคนมารวมกันทุกข้อแล้วหาค่าเฉลี่ย (average or mean scale value) โดยหารผลรวมด้วยจำนวนข้อทั้งหมด ค่าเฉลี่ยที่ได้นี้ก็มิได้มีความหมายเหมือนกับค่าเฉลี่ยในมาตราอันตรภาค แต่ถือว่าอยู่ในมาตราจัดอันดับเท่านั้น

จากข้อเสนอแนะดังกล่าว สรุปได้ว่า การเลือกใช้สถิติให้เหมาะสมกับระดับการวัดของข้อมูลนั้นนับว่าเป็นสิ่งสำคัญมาก โดยเฉพาะอย่างยิ่งข้อมูลที่จัดอยู่ในมาตราจัดอันดับซึ่งปรากฏอยู่มากในงานวิจัยทางด้านพฤติกรรมศาสตร์ และสถิติทดสอบที่เหมาะสมกับข้อมูลประเภทนี้ก็คือ สถิติทดสอบประเภทนันพาราเมตริก ซึ่งวิธีการของสถิติทดสอบเหล่านี้ได้รับการพัฒนาเพื่อให้สามารถใช้กับข้อมูลที่อยู่ในสถานการณ์ต่าง ๆ ได้อย่างกว้างขวาง เช่น เมื่อข้อมูลอยู่ในมาตราจัดอันดับก็สามารถใช้สถิติทดสอบต่อไปนี้ในการทดสอบได้เช่นกัน คือ วิลค็อกซอน เทส (Wilcoxon test) ใช้ในการทดสอบซึ่งคล้ายคลึงกับ ที เทส (t-test) เอช เทส ของคราสคัล วอลลิส (The Kruskal Wallis'H-test) ใช้ในการทดสอบซึ่งคล้ายคลึงกับ เอฟ เทส (F-test) ในการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) และ ฟรีดแมน เทส (Friedman test) ใช้ในการทดสอบซึ่งคล้ายคลึงกับเอฟ เทส (F-test) ในการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) (Derek Srisukho, 1974)

ในงานวิจัยทางพฤติกรรมศาสตร์ที่ใช้ระเบียบวิธีวิจัยเชิงทดลองนั้น วิธีการวิเคราะห์ข้อมูลอย่างหนึ่งที่นิยมใช้กันทั่วไปคือ การวิเคราะห์ความแปรปรวน (Analysis of Variance) ซึ่งมีลักษณะแตกต่างกันไปได้หลายรูปแบบ ขึ้นอยู่กับการออกแบบการวิจัยเป็นสำคัญ สำหรับวิธีวิเคราะห์ทางสถิติที่ใช้ในการวิจัยที่ง่ายที่สุด ก็คือ การวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ซึ่งการแยกความแปรปรวนทั้งหมดที่เกิดขึ้นในการทดลองนั้น ผู้วิจัยได้แยกเอาความแปรปรวนอันเนื่องจากสิ่งทดลองเท่านั้นออกมา และส่วนที่เหลือจะเป็นความคลาดเคลื่อน ดังนั้น ผลการวิเคราะห์ข้อมูลจะถูกต้อง ใกล้เคียงกับความเป็นจริง ก็ต่อเมื่อ

กลุ่มทดลองและกลุ่มควบคุมต้องมีลักษณะที่คล้ายคลึงกัน (homogeneous) แต่อย่างไรก็ตาม ในทางปฏิบัติแล้วผู้วิจัยมักจะประสบปัญหาที่เนื่องมาจากความแตกต่างของหน่วยทดลอง และถ้ามีความแตกต่างทางด้านหน่วยทดลองเกิดขึ้นแล้ว ความแปรปรวนที่เกิดจากหน่วยทดลองก็จะไปรวมอยู่กับความคลาดเคลื่อน ซึ่งจะทำให้ความคลาดเคลื่อนมีมากขึ้น ดังนั้นเพื่อขจัดความแปรปรวนที่เนื่องมาจากหน่วยทดลองออกไปจากความคลาดเคลื่อน จึงควรจัดหน่วยทดลองที่เหมือนกัน หรือคล้าย ๆ กัน ให้อยู่ในกลุ่ม (block) เดียวกัน ซึ่งก็คือการเลือกใช้แบบแผนการทดลองที่เรียกว่าการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม (Randomized Block Design)

จากหลักการดังกล่าว จะพบว่า รูปแบบการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่มนี้จะช่วยขจัดความแปรปรวนที่เกิดจากความแตกต่างของหน่วยทดลองลงได้ ซึ่งจะเป็นผลทำให้การศึกษาอิทธิพลของสิ่งทดลองต่างๆ เด่นชัดขึ้น อีกทั้งยังส่งผลให้การวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่มมีความแม่นยำ (precision) ในการทดสอบสมมติฐาน สูงกว่าการวิเคราะห์ความแปรปรวนแบบทางเดียว (สุวัฒนา สุวรรณเขตนิคม, 2527)

ด้วยคุณสมบัติที่ดีกว่า ของการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่มดังกล่าว จึงทำให้แบบแผนการทดลองนี้ เป็นที่นิยมใช้กันมากขึ้น สำหรับงานวิจัยเชิงทดลองในปัจจุบัน แต่อย่างไรก็ตามปัญหาสำคัญอีกประการหนึ่ง ที่ผู้วิจัยมักจะประสบอยู่ก็คือ ตัวแปรบางตัวที่ผู้วิจัยต้องการศึกษานั้น ไม่สามารถวัดออกมาถึงมาตราอันตรภาคที่แท้จริงได้ เช่น ความสามารถทางด้านการวาดภาพ ดนตรี กีฬา การคัดลายมือ การเขียนเรียงความ งานทางด้านศิลปะและงานฝีมืออื่น ๆ ตัวแปรดังกล่าวจะปรากฏอยู่มากในงานวิจัยทางด้านการศึกษา โดยที่การวัดตัวแปรเหล่านี้ จะนิยมให้ผู้เชี่ยวชาญเป็นผู้พิจารณา และประเมินค่าตัวแปรข้างต้นซึ่งผู้เชี่ยวชาญก็มักจะประเมินค่าตัวแปรออกมาเป็นคะแนน แต่อย่างไรก็ตาม เมื่อพิจารณาถึงลักษณะของคะแนนที่ได้จากตัวแปรประเภทดังกล่าวแล้ว จะเห็นว่าคะแนนเหล่านี้วัดได้แค่ระดับอันดับเท่านั้น ด้วยเหตุนี้จึงมีผู้วิจัยเป็นจำนวนมากที่ไม่กล้านำข้อมูลเหล่านั้น

ไปวิเคราะห์แล้วทดสอบสมมติฐานด้วยสถิติทดสอบเอฟ (F-test) ในการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่มทั้งนี้เพราะ สถิติทดสอบเอฟ จะใช้ทดสอบความแตกต่างของค่าเฉลี่ยของประชากร ($H_0 : \mu_1 = \mu_2 \ldots = \mu_k$) ซึ่งค่าเฉลี่ยจะเป็นค่ากลางที่มีความหมายก็ต่อเมื่อข้อมูลที่นำมาวิเคราะห์จัดอยู่ในระดับการวัดมาตราอันตรภาค หรือมาตราอัตราส่วนเท่านั้น ส่วนข้อมูลที่อยู่ในระดับการวัดมาตราจัดอันดับแล้ว ค่ากลางที่ใช้ควรจะเป็นค่ามัธยฐาน (median)

ฮอดเจส และ เลย์แมน (Hodges and Lehmann, 1962) ได้เสนอสถิติทดสอบรูปแบบหนึ่ง ซึ่งนำมาใช้วิเคราะห์ข้อมูลในลักษณะเดียวกับสถิติทดสอบเอฟในการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม (Randomized Block Design) โดยที่สถิติทดสอบที่เสนอขึ้นนี้ จะนำไปใช้วิเคราะห์ข้อมูลที่อยู่ในรูปของอันดับ (rank) ซึ่งต่อมาสถิติทดสอบนี้จึงได้ชื่อว่าสถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท (Hodges-Lehmann Alignment test)

สถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท ใช้ทดสอบสมมติฐานเพื่อเปรียบเทียบอันดับของประชากรตั้งแต่ 2 กลุ่มขึ้นไปที่สัมพันธ์กัน ซึ่งได้รับการปรับปรุงมาจากรูปแบบของวิลค็อกซอน (Combined Wilcoxon model) โดยที่ผู้วิจัยจะต้องถ่วงน้ำหนักของข้อมูลเพื่อขจัดความแตกต่างระหว่างบล็อคเสียก่อน ซึ่งวิธีการถ่วงน้ำหนักข้อมูลนั้นก็จะใช้ค่าประมาณตำแหน่งของประชากร (estimate of the location of population) อันได้แก่ ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน ค่าเฉลี่ยทริมด์ (trimmed mean) หรือ ค่าเฉลี่ยวินโซไรซด์ (Winsorized mean) ของแต่ละบล็อคไปลบออกจากข้อมูลเดิม จากนั้นจึงนำข้อมูลที่ได้ในทุกๆ บล็อคมารวมกันเพื่อจัดเรียงอันดับจากน้อยไปมาก แล้วทดสอบสมมติฐานต่อไป ซึ่งจากการศึกษาของ ฮอดเจส (Hodges) และเลย์แมน (Lehmann) พบว่า ประสิทธิภาพของสถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท โดยใช้วิธีการถ่วงน้ำหนักข้อมูลด้วยค่าเฉลี่ยเลขคณิตของแต่ละบล็อค และจำนวนตัวแปรทดลอง (treatment) เท่ากับ 2 จะมีประสิทธิภาพไม่ต่ำกว่า 0.955 เมื่อเปรียบเทียบกับสถิติทดสอบเอฟ (F-test) (Marascuilo, 1977)

จากการศึกษาวิธีการถ่วงน้ำหนักข้อมูลของสถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท พบว่า นอกจากการถ่วงน้ำหนักข้อมูลด้วยค่าเฉลี่ยเลขคณิตแล้ว ก็ยังมีค่ามัธยฐาน ค่าเฉลี่ยทริมด์ และค่าเฉลี่ยวินโซไรซด์ ที่สามารถนำมาถ่วงน้ำหนักข้อมูลได้เช่นเดียวกันและ ขั้นตอนของการวิเคราะห์ข้อมูลก็ยังคงเหมือนกับ การถ่วงน้ำหนักข้อมูลโดย วิธีการใช้ค่าเฉลี่ยเลขคณิต แต่อย่างไรก็ตาม เมื่อพิจารณาจากข้อมูลชุดเดียวกันแล้ว จะพบว่า ค่าประมาณทั้ง 3 ชนิดนี้ เป็นค่าที่แตกต่างกัน กล่าวคือ ค่ามัธยฐานเป็นค่ากลางที่แบ่งข้อมูลเป็น 2 กลุ่มเท่าๆ กัน ตามเกณฑ์ความมาก-น้อย ส่วนค่าเฉลี่ยทริมด์เป็นค่าเฉลี่ยของข้อมูลที่ได้หลังจากการตัดข้อมูล ที่ต่ำกว่าควอไทล์ที่ 1 และสูงกว่าควอไทล์ที่ 3 ออกไปแล้วส่วนค่าเฉลี่ยวินโซไรซด์เป็นค่าเฉลี่ย ของข้อมูลที่ได้จากการแทนค่าข้อมูลที่ต่ำกว่าควอไทล์ที่ 1 ด้วยค่าของควอไทล์ที่ 1 และแทนค่า ข้อมูลที่สูงกว่าควอไทล์ที่ 3 ด้วยค่าของควอไทล์ที่ 3 เพราะฉะนั้น จากตัวอย่างข้อมูลที่กำหนด ให้จะสามารถหาค่าประมาณทั้ง 3 ชนิดได้ ดังต่อไปนี้

60 65 67 74 82 83 84 87 96

ควอไทล์ที่ 1 ควอไทล์ที่ 3

ค่ามัธยฐาน = 82

ค่าเฉลี่ยทริมด์ = (67+74+82+83+84)/5 = 78

ค่าเฉลี่ยวินโซไรซด์ = (67+67+67+74+82+83+84+84+84)/9 = 76.9

เมื่อนำค่าประมาณทั้ง 3 ชนิด มาถ่วงน้ำหนักข้อมูลแล้ว จะทำให้อันดับของข้อมูลที่ ได้แตกต่างกันไป และเมื่อนำอันดับของข้อมูลที่ได้จากการถ่วงน้ำหนักมาเปรียบเทียบกันแล้วก็ จะได้อันดับที่แตกต่างกันด้วย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ด้วยเหตุผลดังกล่าวข้างต้น จึงทำให้ผู้ที่จะนำสถิติทดสอบ ฮอดเจส-เลย์แมน อะไลนเม็นท์ ไปใช้ เกิดความไม่แน่ใจว่า จะเลือกใช้วิธีการถ่วงน้ำหนักข้อมูลด้วยวิธีใดจึงจะเหมาะสม กล่าวคือ วิธีการถ่วงน้ำหนักข้อมูลด้วย ค่ามัธยฐาน ค่าเฉลี่ยทริมด์ หรือ ค่าเฉลี่ยวินโซไรซด์ วิธีใดที่จะทำให้ สถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท มีความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ตามที่ระบุได้ ดังที่ Neyman (1950, อ้างถึงใน Derek Srisukho, 1974) กล่าวว่า เมื่อต้องการเลือกใช้สถิติทดสอบเราต้องเริ่มด้วยการพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 แล้วจึงพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 2 ต่อไป โดยมีลำดับขั้นคือ ให้ความน่าจะเป็นที่ยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน ที่กำหนดไว้ ซึ่งวิธีการศึกษาเพื่อให้ได้ผลสรุปดังกล่าว สามารถศึกษาได้โดยวิธีการซิมูเลชั่น (Simulation) ด้วยเครื่องคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo) ซึ่งเป็นวิธีที่สามารถนำมาดำเนินการเพื่อหาผลสรุปในการทดลองแก้ปัญหาได้ในปัจจุบัน

<u>วัตถุประสงค์ของการวิจัย</u>

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบค่าความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบ ฮอดเจส-เลย์แมน อะไลนเม็นท (Hodges-Lehmann Alignment test) เมื่อใช้วิธีการถ่วงน้ำหนักข้อมูลด้วย ค่ามัธยฐาน (median) ค่าเฉลี่ยทริมด์ (trimmed mean) และ ค่าเฉลี่ยวินโซไรซด์ (Winsorized mean)

<u>สมมติฐานของการวิจัย</u>

การถ่วงน้ำหนักข้อมูลด้วยค่ามัธยฐาน ค่าเฉลี่ยทริมด์ และค่าเฉลี่ยวินโซไรซด์ ทำให้สถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท สามารถควบคุมค่าความคลาดเคลื่อนประเภทที่ 1 ได้เท่ากับระดับความมีนัยสำคัญที่กำหนด

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ศึกษาตัวแปรดังต่อไปนี้

   1.1 ตัวแปรตาม (Dependent Variable) คือค่าความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบฮอตเจส-เลย์แมน อะไลนเม็นท

   1.2 ตัวแปรอิสระ (Independent Variable) คือ ค่าประมาณที่ใช้ในการถ่วงน้ำหนักข้อมูล ได้แก่ ค่ามัธยฐาน ค่าเฉลี่ยทริมด์ และค่าเฉลี่ยวินโซไรซด์

2. ศึกษาเฉพาะการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม เมื่อ

   2.1 แผนการวิจัยเชิงทดลองเป็นแบบ Nonrepeated-measures design

   2.2 จำนวนบล็อค (block) เท่ากับ 2

   2.3 จำนวนตัวแปรทดลอง (treatment) เท่ากับ 3

   2.4 กลุ่มตัวอย่างมีขนาด 30,60 และ 90

   2.5 โมเดลของการวิเคราะห์เป็นแบบไม่มีปฏิสัมพันธ์ (interaction) ระหว่างตัวแปรทดลอง กับ บล็อค

3. ศึกษาค่าความคลาดเคลื่อนประเภทที่ 1 จากประชากร ที่มีลักษณะการแจกแจงเป็นแบบปกติเท่านั้น

4. กำหนดพารามิเตอร์ $\mu = 500$ และ $\sigma^2 = 100$ สำหรับการทดลองเมื่อการแจกแจงเป็นแบบปกติ

5. การวิจัยครั้งนี้กำหนดขนาดความคลาดเคลื่อนที่ระบุ (α) 2 ระดับ คือ $\alpha = .05$ และ $\alpha = .01$

6. กำหนดให้ระดับของความแตกต่างกันของผลที่ได้รับจากการที่อยู่ในบล็อคของประชากรเป็นเบต้า ($\beta_{b..}$)

จากโมเดลการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม เมื่อไม่มีปฏิสัมพันธ์ระหว่างตัวแปรบล็อคกับตัวแปรทดลอง

$$X_{ikb} = \mu + \alpha_k + \beta_{b.} + \ell_{ikb} \qquad \begin{matrix} i = 1, 2, \ldots, n \\ k = 1, 2, \ldots, K \\ b = 1, 2, \ldots, B \end{matrix}$$

เนื่องจากการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่มจะมีความแม่นยำในการทดสอบสมมติฐาน มากขึ้น เมื่อตัวแปรบล็อคที่นำมาใช้ในการวิจัยนั้นมีความสัมพันธ์กับตัวแปรตามที่ต้องการศึกษา ดังนั้น ในการวิจัยครั้งนี้ จึงกำหนดให้ตัวแปรบล็อคมีความสัมพันธ์กับตัวแปรตามในระดับปานกลาง คือ ประมาณ 0.4 ถึง 0.6 ซึ่งจะได้ค่าสัดส่วนของความแปรผันร่วมกันระหว่างตัวแปรทั้งสอง ประมาณ 16% ถึง 36% และเมื่อเปิดตารางพื้นที่ภายใต้โค้งปกติแล้วจะได้ระดับของความแตกต่างกันของผลที่ได้รับจากการที่อยู่ ในบล็อคของประชากร ($\beta_{b.}$) มีค่าทั้งหมด 9 ค่าดังนี้ 0.46,0.56,0.66,0.76,0.86,0.96,1.06,1.16และ 1.26

7. แต่ละกรณีของการวิจัยครั้งนี้จะทำการทดลองซ้ำจำนวน 4,000 ครั้ง

8. การวิจัยนี้จำลองการทดลองขึ้นในคอมพิวเตอร์โดยอาศัยเทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation Technique) และ โปรแกรมสัปรูทีนโดยจำกัดการศึกษาเฉพาะลักษณะการแจกแจงของประชากรเป็นแบบปกติ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

คำจำกัดความ

1. ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐานศูนย์ เมื่อสมมติฐานศูนย์เป็นจริง โอกาสที่จะทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 กำหนดด้วย

2. ค่าความคลาดเคลื่อนประเภทที่ 1 จากการทดลอง (Actual Type I Error) หมายถึง ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ที่เกิดขึ้นจริงในการทดสอบสมมติฐานของการทดสอบตามแผนการทดลองครั้งนี้ ซึ่งหาได้จากสัดส่วนของจำนวนครั้งที่ปฏิเสธสมมติฐานศูนย์เมื่อสมมติฐานศูนย์นั้นเป็นจริง กับจำนวนครั้ง ของการทดลองทั้งหมด 4,000 ครั้ง

3. ขนาดความคลาดเคลื่อนที่ระบุ (Nominated α) หมายถึง ขนาดความคลาดเคลื่อนประเภทที่ 1 ที่ควบคุมด้วยระดับนัยสำคัญ (α) ซึ่งผู้วิจัยเป็นผู้กำหนดขึ้น 2 ขนาด คือ $\alpha = .05$ และ $\alpha = .01$

4. การถ่วงน้ำหนักข้อมูล (Alignment) หมายถึง การนำเอาค่าประมาณตำแหน่งของประชากรบางค่า (estimate of the location of population) ไปลบออกจากข้อมูลดิบของแต่ละบล็อค โดยที่ค่าประมาณตำแหน่งของประชากรดังกล่าวจะต้องมีความสัมพันธ์แบบสมมาตรกับข้อมูลดิบ

5. มัธยฐาน หมายถึง ค่ากลางที่แบ่งข้อมูลเป็น 2 กลุ่มเท่า ๆ กัน ตามเกณฑ์ความมาก-น้อย

6. ค่าเฉลี่ยทริมด์ (trimmed mean) หมายถึง ค่าเฉลี่ยของข้อมูลที่ได้หลังจากการขจัดข้อมูลที่ต่ำกว่าควอไทล์ที่ 1 และสูงกว่าควอไทล์ที่ 3 ออกไปแล้ว

7. ค่าเฉลี่ยวินโซไรซด์ (Winsorized mean) หมายถึง ค่าเฉลี่ยของข้อมูลที่ได้จากการแทนค่าข้อมูลที่ต่ำกว่าควอไทล์ที่ 1 ด้วยค่าของควอไทล์ที่ 1 และแทนค่าข้อมูลที่สูงกว่าควอไทล์ที่ 3 ด้วยค่าของควอไทล์ที่ 3

8. เบต้า ($\beta_{b..}$) หมายถึง ระดับของความแตกต่างกัน ของผลที่ได้รับ จากการที่อยู่ในบล็อคของประชากร ซึ่งมีวิธีการคำนวณดังกล่าวไว้ใน ขอบเขตของการวิจัย

9. สถิติทดสอบฮอดเจส-เลย์แมน อะไลนเม็นท (Hodges-Lehmann Alignment Test) หมายถึง สถิติที่ใช้ทดสอบสมมติฐาน เพื่อเปรียบเทียบอันดับของประชากรตั้งแต่ 2 กลุ่มขึ้นไป โดยจะต้องถ่วงน้ำหนักข้อมูล ด้วยค่าประมาณตำแหน่งของประชากรก่อนแล้วจึงนำข้อมูลที่ได้มาจัดเรียงอันดับ

ประโยชน์ที่คาดว่าจะได้รับ

การวิจัยครั้งนี้ให้ประโยชน์ในการหาหลักฐานและผลสรุป ในการที่จะแนะนำผู้ใช้การวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม และทดสอบสมมติฐาน ด้วยสถิติทดสอบฮอดเจส–เลย์แมน อะไลนเม็นท สามารถมีเกณฑ์ในการตัดสินใจว่า ควรจะใช้วิธีการใดในการถ่วงน้ำหนักข้อมูลก่อนที่จะนำข้อมูลเหล่านั้นไปวิเคราะห์ ซึ่งจะทำให้ผลของการตัดสินใจมีประสิทธิภาพ และเหมาะสมกับสถานการณ์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย