

การวิจัยเชิงทดลอง (Experimental Research)

การวิจัยเชิงทดลอง เป็นกระบวนการศึกษาหาความจริงเกี่ยวกับตัวแปร โดยที่ผู้วิจัยสามารถจัดการกระทำตัวแปรต้น โดยจัดให้เกิดความแปรผันหรือความแปรปรวนของตัวแปรต้นที่ต้องการศึกษาให้มีค่ามากน้อยตามความต้องการได้อย่างเหมาะสม และในขณะที่เดียวกันก็ต้องควบคุมตัวแปรแทรกซ้อน (extraneous variable) ไม่ให้มามีผลกระทบต่อผลการทดลอง

ในการวิจัยเชิงทดลองแต่ละครั้ง ผู้วิจัยจะต้องดำเนินการออกแบบการวิจัย (research design) ให้เหมาะสมกับปัญหาที่ต้องการศึกษา ซึ่งแบบการวิจัยที่เหมาะสมนี้จะช่วยให้ได้ผลการทดลองที่ถูกต้องสามารถตอบคำถามการวิจัยและทดสอบสมมติฐานได้ตามวัตถุประสงค์ที่ตั้งไว้ นอกจากนี้แบบการวิจัยที่ดียังจะช่วยป้องกันและลดปัจจัยภายนอกที่จะเข้ามามีผลกระทบต่อผลการทดลองได้อย่างมาก อันจะส่งผลให้การวิจัยนั้นๆ มีความตรงของการวิจัยสูง ซึ่งความตรงของการวิจัยสามารถแบ่งได้เป็น 2 ลักษณะ คือ ความตรงภายใน (internal validity) และความตรงภายนอก (external validity)

ความตรงภายในเป็นคุณสมบัติของงานวิจัยที่สามารถสรุปได้ว่าผลการวิจัยที่ได้นั้นเป็นผลมาจากสิ่งทดลอง (treatment) หรือตัวแปรต้นที่ผู้วิจัยจัดการกระทำขึ้นในการทดลองเท่านั้น ไม่ได้เกิดจากปัจจัยอื่นๆ ร่วมด้วย ส่วนความตรงภายนอกนั้น เป็นคุณสมบัติของงานวิจัยที่สามารถขยายการสรุปผลการวิจัย (generalization) ออกไปให้ครอบคลุมในกลุ่มประชากรที่ใหญ่กว่าได้ ดังนั้น ในการออกแบบการวิจัยทุกๆ ครั้ง ผู้วิจัยจะต้องให้ความสำคัญกับความตรงทั้งสองประเภทของการวิจัย โดยเฉพาะอย่างยิ่งความตรงภายใน ซึ่งจะต้องนำมาพิจารณาเป็นอันดับแรก และปัจจัยหนึ่งที่มีผลกระทบเป็นอย่างมากต่อความตรงภายในของการวิจัย ก็คือ ความแปรปรวนที่เกิดจากตัวแปรแทรกซ้อน

ตัวแปรแทรกซ้อน เป็นตัวแปรที่ผู้วิจัยไม่ต้องการศึกษาแต่จะเข้ามาบรบกวนและส่งผลต่อตัวแปรอิสระที่ศึกษาทำให้ผลการวัดค่าตัวแปรตามคลาดเคลื่อนไปได้ ดังนั้นในการดำเนินการวิจัยทุกครั้งผู้วิจัยจึงต้องทำการควบคุมตัวแปรแทรกซ้อนให้เกิดขึ้นน้อยที่สุด ซึ่งวิธีการควบคุมตัวแปรประเภทนี้สามารถทำได้โดย

1. การกำจัดตัวแปรแทรกซ้อนออกไปจากการวิจัย (elimination) โดยการควบคุมให้ความแปรปรวนของตัวแปรแทรกซ้อนมีค่าต่ำสุดหรือมีค่าใกล้เคียงศูนย์มากที่สุดในกลุ่มตัวอย่างที่นำมาศึกษา เช่นถ้าคาดว่าเพศเกี่ยวข้องกับผลการทดลอง ในการทำการทดลองก็เลือกกลุ่มตัวอย่างที่เป็นเพศเดียวกันทั้งหมดมาทำการทดลอง แต่อย่างไรก็ตาม การควบคุมตัวแปรแทรกซ้อนด้วยวิธีนี้เป็นการลดการขยายผลการวิจัย (generalization) เพราะผู้วิจัยไม่สามารถจะลงสรุปความไปยังกลุ่มประชากรทั้งหมดได้ จะลงสรุปความเฉพาะเพศที่นำมาศึกษา เท่านั้น
2. การทำให้ตัวแปรแทรกซ้อนมีค่าคงที่ตลอดการทดลอง (keeping it constant) โดยการจัดให้ตัวแปรแทรกซ้อนมีลักษณะคงที่ หรือเหมือนกันทุกครั้งที่ในการศึกษาแต่ละกลุ่มตัวอย่าง เช่น ถ้าอุณหภูมิเป็นตัวแปรแทรกซ้อนในการศึกษาก็จะจัดให้หน่วยทดลองอยู่ในสภาพอากาศที่ปรับให้อุณหภูมิเท่ากันตลอดเวลาที่ศึกษา
3. การถ่วงดุล (balanceing) โดยการทำให้อิทธิพลของตัวแปรแทรกซ้อน มีค่าเท่าเทียมกันด้วยการจับคู่ (matching) คือจัดหน่วยทดลองที่มีค่าตัวแปรนั้นๆ เท่ากันเป็นคู่แล้วแยกคนหนึ่งไว้ในกลุ่มทดลอง อีกคนหนึ่งไว้ในกลุ่มควบคุม วิธีการจับคู่กระทำได้ไม่ยากในการที่ตัวแปรแทรกซ้อนที่ต้องการควบคุมมีจำนวนไม่มากนัก แต่ถ้ามีมากเกินไปก็จะมีวิธีการถ่วงดุลอีกวิธีหนึ่ง ซึ่งเป็นวิธีที่สะดวก มีประสิทธิภาพสูง และสามารถควบคุมตัวแปรแทรกซ้อนได้หลายๆ ตัวพร้อมกัน ก็คือ การสุ่ม (randomization) ด้วยวิธีการสุ่มจะทำให้ค่าตัวแปรที่แตกต่างกันของหน่วยทดลองในกลุ่มต่างๆ จะชดเชยกัน เมื่อเฉลี่ยแล้วค่าตัวแปรเหล่านั้นในกลุ่มต่างๆ จะมีค่าใกล้เคียงกันได้ ซึ่งการสุ่มอาจทำได้ทั้งการแบ่งกลุ่มโดยการสุ่มและการให้ปัจจัยทดลองแก่กลุ่มต่างๆ อย่างสุ่ม (random assignment)

4. การจัดตัวแปรแทรกซ้อนให้เป็นตัวแปรอิสระอีกตัวหนึ่ง (built into the design) แล้วใช้แบบการวิจัยแบบแฟคทอเรียล (factorial) แบบการวิจัยประเภทนี้ จะช่วยให้ผู้วิจัยสามารถแยกความแปรปรวนที่เนื่องมาจากตัวแปรแทรกซ้อน ออกจากความแปรปรวนทั้งหมดได้และยังสามารถศึกษาถึงปฏิสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรแทรกซ้อนได้ นอกจากนี้แบบการวิจัยแบบแฟคทอเรียล ยังสามารถลดความแปรปรวนเนื่องจากความคลาดเคลื่อนให้น้อยลงได้ด้วย

5. การควบคุมโดยวิธีการทางสถิติ (statistical control) โดยการวัดค่าตัวแปรแทรกซ้อนที่ต้องการควบคุมออกมาเป็นตัวเลข แล้วใช้เทคนิควิธีทางสถิติที่สามารถขจัดอิทธิพลของตัวแปรแทรกซ้อนนี้วิเคราะห์ข้อมูล เช่น การหา partial correlation หรือการวิเคราะห์ความแปรปรวนร่วม เป็นต้น

จากวิธีการควบคุมตัวแปรแทรกซ้อน ที่กล่าวมาข้างต้น สรุปได้ว่าวิธีการโดยทั่วไปที่ใช้ ในการควบคุมตัวแปรแทรกซ้อน แบ่งออกเป็น 3 วิธี ดังนี้

1. การควบคุมที่ได้มาโดยการเลือกใช้แบบการวิจัยที่เหมาะสม ทั้งนี้เพราะจุดประสงค์ประการหนึ่งของการออกแบบการวิจัย คือ เพื่อกำจัดอิทธิพลที่แตกต่างกันของตัวแปรแทรกซ้อน
2. การควบคุมโดยใช้วิธีการทางสถิติ
3. การนำวิธีการควบคุมตัวแปรแทรกซ้อนหลายๆ วิธีมาใช้ร่วมกันในแบบการวิจัยหนึ่งๆ

การวิเคราะห์ข้อมูลที่ใช้ในการวิจัยเชิงทดลอง

เมื่อเก็บรวบรวมข้อมูลจากการทดลองได้แล้ว ผู้วิจัยจะนำข้อมูลเหล่านั้นไปวิเคราะห์เพื่อสรุปผลการทดลอง โดยพิจารณาจากผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร ถ้าเป็นการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากร

สองกลุ่ม สถิติทดสอบที่นำมาใช้ทดสอบสมมติฐาน คือ z-test หรือ t-test แต่ถ้าผู้วิจัยต้องการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากรที่มากกว่าสองกลุ่มขึ้นไป ก็จะใช้วิธีการวิเคราะห์ข้อมูลที่เรียกว่าการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA)

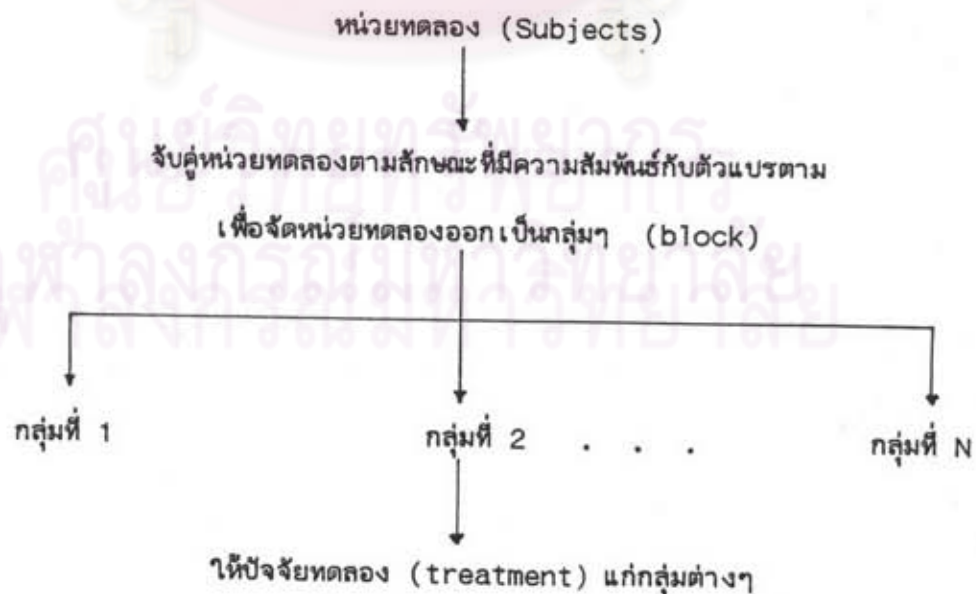
การวิเคราะห์ความแปรปรวนเป็นวิธีการวิเคราะห์ข้อมูลอย่างหนึ่งที่น่าสนใจกันทั่วไป ในทางวิจัยเชิงทดลองโดยสามารถประยุกต์ใช้ได้หลายแบบด้วยกันขึ้นอยู่กับการออกแบบการทดลองเป็นสำคัญ สำหรับรูปแบบการวิเคราะห์ความแปรปรวนที่ง่ายที่สุด คือ การวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) โดยผู้วิจัยจะแยกความแปรปรวนทั้งหมดที่เกิดขึ้นในการทดลองออกเป็น 2 ส่วน คือ ความแปรปรวนระหว่างกลุ่ม (between-group variance) ซึ่งเป็นความแปรปรวนอันเนื่องมาจากสิ่งทดลอง (treatment) และส่วนที่เหลือคือความแปรปรวนภายในกลุ่ม (within-group variance) หรือความคลาดเคลื่อน (error) ซึ่งปัจจัยสำคัญอย่างหนึ่งที่ทำให้เกิดความแปรปรวนชนิดนี้ก็คือ ความแตกต่างระหว่างหน่วยทดลอง (subject) จากแนวคิดในการแยกความแปรปรวนดังกล่าวจะนำมาใช้เป็นหลักสำคัญในการทดสอบสมมติฐานของการวิเคราะห์ความแปรปรวนโดยใช้อัตราส่วนของความแปรปรวนระหว่างกลุ่มกับความแปรปรวนภายในกลุ่มหรือความคลาดเคลื่อนเป็นหลักในการสรุปผลการวิจัยและ เรียกอัตราส่วนนี้ว่าอัตราส่วนเอฟ (F-ratio) ถ้าอัตราส่วนเอฟมีค่ามากโอกาสที่จะปฏิเสธสมมติฐานก็จะมีมาก หรืออาจกล่าวได้ว่า ถ้าความแปรปรวนระหว่างกลุ่มมีมากโอกาสที่ค่าเฉลี่ยของแต่ละกลุ่มจะแตกต่างกันก็มีมากด้วย

การทำให้ความแปรปรวนระหว่างกลุ่มหรือความแปรปรวนของตัวแปรตามมีค่าสูงสุดถือว่าเป็นหลักสำคัญในการออกแบบการวิจัยประการหนึ่ง ที่ผู้วิจัยจะต้องนำมาพิจารณา โดยที่ในการวิจัยแต่ละครั้งผู้วิจัยจะต้องพยายามจัดสิ่งทดลองให้มีความแตกต่างกันมากที่สุด เพื่อให้เห็นผลของการเปรียบเทียบตัวแปรตามได้อย่างชัดเจนที่สุด แต่อย่างไรก็ตามในทางปฏิบัติแล้ว ผู้วิจัยมักจะประสบปัญหาที่ว่าความแปรปรวนของตัวแปรตามนั้นมีส่วนหนึ่งที่

เกิดมาจากตัวแปรแทรกซ้อน โดยเฉพาะอย่างยิ่งในกรณีที่ตัวแปรแทรกซ้อนนั้นๆ เป็นตัวแปรที่มีลักษณะซึ่งผู้วิจัยจัดการกระทำไม่ได้ เช่น I.Q., ความรู้เดิมของหน่วยทดลอง เป็นต้น ซึ่งวิธีการหนึ่งที่จะช่วยลดความแปรปรวนที่เกิดจากตัวแปรแทรกซ้อน ก็คือ ผู้วิจัยจะต้องเลือกหน่วยทดลองที่มีลักษณะคล้ายคลึงกันให้มากที่สุด นำมาจัดให้อยู่ในกลุ่มเดียวกัน แล้วใช้แบบแผนการทดลองที่เรียกว่าการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม (Randomized Block Design)

การวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม เป็นแบบแผนการทดลองที่นำหลักการของการจับคู่ (matching) มาใช้ในการควบคุมตัวแปรแทรกซ้อน โดยก่อน การทดลอง ผู้วิจัยจะแบ่งหน่วยทดลองออกเป็นกลุ่มๆ (block) ซึ่งหน่วยทดลองที่อยู่ภายในกลุ่มเดียวกัน จะมีลักษณะคล้ายคลึงกันมากที่สุด โดยที่ลักษณะ ที่คล้ายคลึงกันของหน่วยทดลองนี้ จะต้องมีความสัมพันธ์กับตัวแปรตามหลังจากแบ่งหน่วยทดลองออกเป็นกลุ่มๆ แล้ว ผู้วิจัยก็จะให้ปัจจัยทดลอง (treatment) แก่กลุ่มต่างๆ

จากที่กล่าวมาข้างต้น สามารถสรุปแบบแผนการทดลองได้ดังนี้





โมเดลสำหรับการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม

1. โมเดลการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม เมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่างตัวแปรทดลอง (treatment) กับตัวแปรบล็อก (block) มีรูปแบบดังนี้

$$X_{bki} = \mu + \alpha_k + \beta_{b..} + \epsilon_{bki}$$

- เมื่อ X_{bki} คือ ผลการวัดตัวแปรตามของตัวอย่างที่สุ่มมาจากประชากร
- μ คือ มีชดิมเลขคณิตของประชากร
- α_k คือ ผลจากการทดลอง (treatment effect)
- $\beta_{b..}$ คือ ผลจากการที่อยู่ในบล็อก (block effect)
- ϵ_{bki} คือ ความคลาดเคลื่อนของการทดลอง

2. ข้อตกลงเบื้องต้น

1. ประชากรมีการแจกแจงแบบปกติ (normally distributed population)
2. ความเป็นเอกพันธ์ของความแปรปรวนของประชากร (homogeneity of population error variances)
3. ความเป็นอิสระในการสุ่มตัวอย่าง (independence within samples)
4. ไม่มีปฏิสัมพันธ์ระหว่างตัวแปรบล็อกกับตัวแปรทดลอง

3. สมมติฐานศูนย์ที่ต้องการทดสอบ คือ ไม่มีความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรผลลัพธ์ X ในประชากรทั้ง K กลุ่ม ซึ่งเขียนเป็นสัญลักษณ์ได้ดังนี้

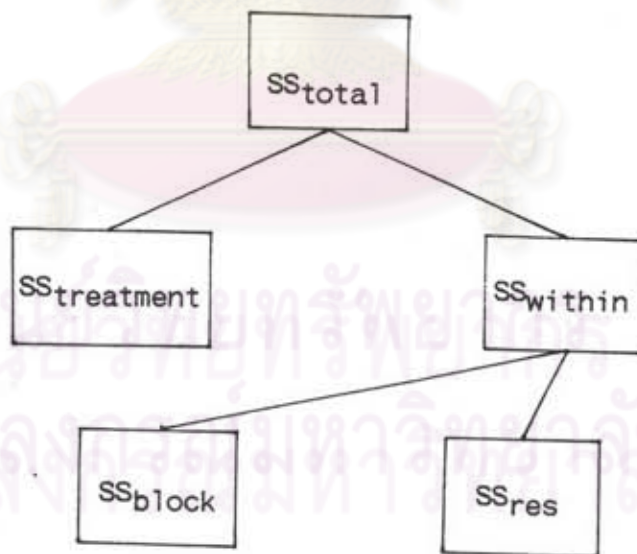
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

4. สถิติทดสอบ โดยใช้อัตราส่วนเอฟ (F-ratio) จากสูตร

$$\text{อัตราส่วนเอฟ} = \frac{\text{ความแปรปรวนระหว่างกลุ่ม}}{\text{ความแปรปรวนภายในกลุ่ม}}$$

$$\text{หรือ } F = \frac{MS_{BG}}{MS_{WG}}$$

จากโมเดลการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม ถ้านำความแปรปรวนทั้งหมด (SS_{total}) มาแจกแจง ก็จะได้ตามแผนผังดังนี้



จากแผนผัง จะเห็นว่าในการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่มนั้น ผู้วิจัยสามารถจำแนกความแปรปรวนภายในกลุ่ม (SS_{within}) ออกได้เป็น 2 ส่วน คือ ความแปรปรวนเนื่องจากกลุ่ม (SS_{block}) และความคลาดเคลื่อนที่เหลือ (SS_{res}) ผลก็คือ ความคลาดเคลื่อนที่เหลือจะมีค่าน้อยลง ซึ่งเมื่อนำความแปรปรวนไปทดสอบด้วยสถิติทดสอบ เอฟ (F-test) แล้วโอกาสที่จะปฏิเสธ สมมติฐานก็จะมีมากขึ้นด้วย

การวิเคราะห์ข้อมูลในงานวิจัยเชิงทดลองที่กล่าวมาทั้งหมดข้างต้นนั้น จะนำมาใช้วิเคราะห์ได้ก็ต่อเมื่อข้อมูลที่เก็บรวบรวมได้จัดอยู่ในระดับการวัดมาตราอันตรภาค (interval scale) หรือมาตราอัตราส่วน (ratio scale) แต่สำหรับการวิจัยเชิงทดลองในปัจจุบันพบว่าตัวแปรบางประเภทที่ผู้วิจัยต้องการศึกษานั้นไม่สามารถวัดออกมาถึงมาตราอันตรภาคที่แท้จริงได้ เช่น ความสามารถทางด้านการวาดภาพ การตัดลายมือ การเขียนความเรียง พัฒนาการทางด้านการดนตรี พลาณามัย งานทางด้านศิลปะตลอดจนถึงงานฝีมืออื่นๆ ตัวแปรดังกล่าวจะปรากฏอยู่มากในงานวิจัยทางการศึกษา ซึ่งผู้เชี่ยวชาญก็มักจะประเมินค่าตัวแปรออกมาเป็นคะแนน แต่อย่างไรก็ตามเมื่อพิจารณาถึงลักษณะของคะแนนที่ได้แล้ว จะเห็นว่าคะแนนเหล่านี้จัดอยู่ในมาตราจัดอันดับเท่านั้น ด้วยเหตุนี้จึงมีผู้วิจัยเป็นจำนวนมากที่ไม่กล้านำข้อมูลในลักษณะ เช่นนี้ไปวิเคราะห์แล้วทดสอบสมมติฐานด้วยสถิติทดสอบเอฟ (F-test) ทั้งนี้เพราะสถิติทดสอบเอฟ จะใช้ทดสอบความแตกต่างของค่าเฉลี่ยของประชากร

($H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$) ซึ่งค่าเฉลี่ยจะเป็นค่ากลางที่มีความหมายก็ต่อเมื่อข้อมูลที่นำมาวิเคราะห์จัดอยู่ในระดับการวัดมาตราอันตรภาคหรือมาตราอัตราส่วนเท่านั้น

Siegel (1956) กล่าวว่าในงานวิจัยทางด้านพฤติกรรมศาสตร์ ข้อมูลส่วนใหญ่จะวัดได้ถึงมาตราจัดอันดับเท่านั้น และสถิติทดสอบที่เหมาะสมซึ่งจะนำไปใช้ในการทดสอบสมมติฐานก็คือ สถิติทดสอบประเภทนินพาราเมตริก ซึ่งวิธีการของสถิติทดสอบเหล่านี้ได้รับการพัฒนาเพื่อให้สามารถใช้กับข้อมูลที่อยู่ในสถานการณ์ต่างๆ ได้อย่างกว้างขวาง เช่น เมื่อข้อมูลอยู่ในมาตราจัดอันดับก็สามารถใช้สถิติทดสอบต่อไปนี้ ในการทดสอบได้เช่นกัน คือ วิกส์ค็อกซอน เทส (Wilcoxon test) ใช้ในการทดสอบซึ่งคล้ายคลึงกับ ที เทส (t-test)

เอชเทสของคราสคอลล วอลลิส (The Kruskal Wallis'H-test) ใช้ในการทดสอบซึ่งคล้ายคลึงกับ เอฟ เทส (F-test) ในการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) และฟริดแมน เทส (Friedman test) ใช้ในการทดสอบซึ่งคล้ายคลึงกับเอฟเทส ในการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) (Derek Srisukho, 1974)

สถิติทดสอบประเภทอันดับ (Rank test)

สถิติทดสอบประเภทอันดับ ได้เริ่มพัฒนาขึ้นครั้งแรกเพื่อนำมาใช้กับงานวิจัยในกรณีที่มีกลุ่มตัวอย่าง 2 กลุ่ม โดยที่กลุ่มตัวอย่างทั้ง 2 กลุ่ม อาจเป็นกลุ่มตัวอย่างที่สัมพันธ์กันหรือกลุ่มตัวอย่างที่เป็นอิสระต่อกันก็ได้ สำหรับงานวิจัยที่ต้องการศึกษากับกลุ่มตัวอย่าง 2 กลุ่มซึ่งสัมพันธ์กันนั้น สถิติทดสอบประเภทอันดับที่ผู้วิจัยนิยมใช้กันมากคือ วิลค็อกซอน เทส

วิลค็อกซอน เทส ได้พัฒนาวิธีการทดสอบมาจาก Sign test โดยที่วิธีการของ Sign test นั้น เมื่อสุ่มตัวอย่างข้อมูลมาเป็นคู่ๆ แล้ว ก็จะพิจารณาเฉพาะทิศทางของความแตกต่างภายในคู่เท่านั้น ในขณะที่วิธีการของ วิลค็อกซอน เทส จะนำเอาขนาดของความแตกต่างของข้อมูลแต่ละคู่มาพิจารณาด้วย จากวิธีการเปรียบเทียบความแตกต่างภายในแต่ละคู่ของข้อมูลดังกล่าว จึงทำให้สถิติทดสอบวิลค็อกซอน เทส มีประสิทธิภาพของการทดสอบสูงกว่าสถิติทดสอบ Sign test

สำหรับงานวิจัยที่ต้องการศึกษากับกลุ่มตัวอย่างมากกว่า 2 กลุ่มขึ้นไป โดยใช้วิธีวิเคราะห์ความแปรปรวนของลำดับที่ของข้อมูล คือ เอช เทส ของคราสคอลล วอลลิส ซึ่งจะใช้วิธีการจัดเรียงลำดับของข้อมูลที่มีอยู่ทั้งหมด N จำนวน โดยให้อันดับที่ 1 กับข้อมูลที่มีค่าต่ำสุดให้อันดับที่ 2 กับข้อมูลที่มีค่าถัดจากค่าต่ำสุด และทำเช่นนี้ต่อไปจนถึงข้อมูลที่มีค่ามากที่สุดก็ให้เป็นอันดับที่ N ส่วนข้อมูลที่มีค่าซ้ำกันก็จะให้อันดับเฉลี่ย หลังจากนั้นจึงหามลรวมของอันดับที่ของข้อมูลแต่ละกลุ่มแล้วจึงคำนวณค่าสถิติ

ในท่านองเดียวกัน สถิติทดสอบประเภทอันดับที่พัฒนาขึ้นมาใช้ในลักษณะเดียวกัน สถิติทดสอบเอฟใน การวิเคราะห์ความแปรปรวนแบบ 2 ทาง คือ ฟริตแมน เทส โดยในขั้นแรกของการวิเคราะห์ก็จะใช้วิธีการจัดอันดับของข้อมูลแยกเป็นอิสระจากกันในแต่ละกลุ่ม และให้ข้อมูลที่มีค่าต่ำสุดเป็นอันดับ 1 ถ้ามีค่าซ้ำกันก็จะใช้ค่าเฉลี่ยของอันดับแทนในตำแหน่งนั้นๆ

เมื่อพิจารณาวิธีการจัดอันดับของข้อมูลโดยแยกเป็นอิสระจากกันในแต่ละกลุ่มแล้วจะเห็นว่า การจัดอันดับของข้อมูลดังกล่าว ยังขาดการคำนึงถึงสาระสำคัญของประการหนึ่งที่ว่า การกระจายของข้อมูลภายในกลุ่มหนึ่งๆ อาจจะไม่คงที่ หรืออาจเป็นไปได้ว่าค่าของผลที่ได้รับจากการอยู่ในกลุ่มหนึ่งๆ อาจไม่เท่ากัน ซึ่งถ้าหากผู้วิจัยจัดอันดับข้อมูลโดยไม่ได้ คำนึงถึงสาระสำคัญดังกล่าว ก็จะทำให้สารสนเทศที่ผู้วิจัยไม่ต้องการศึกษา เข้าไปรวมอยู่ ในการวิเคราะห์ครั้งนั้นด้วยอันจะส่งผลให้ผลการวิจัยที่ได้ไม่ตรงกับปัญหาของการวิจัยอย่างแท้จริง

เพื่อเป็นการขจัดปัญหาเกี่ยวกับค่าของผลจากการที่อยู่ในกลุ่มที่ไม่เท่ากันและพัฒนา สถิติทดสอบประเภทอันดับให้มีประสิทธิภาพของการทดสอบมากขึ้น ฮอดเจสและเลย์แมน (Hodges and Lehmann, 1962) จึงเสนอสถิติทดสอบประเภทอันดับ ซึ่งนำมาใช้วิเคราะห์ข้อมูลในลักษณะเดียวกับสถิติทดสอบเอฟในการวิเคราะห์ความแปรปรวนแบบกลุ่มกลุ่ม โดยที่ฮอดเจสและเลย์แมน ได้เสนอแนะว่า การทดสอบสมมติฐานโดยการใชโมเดลของวิลค็อกซอน (Combined Wilcoxon model) สามารถพัฒนาให้ดีขึ้นได้โดยจะต้องถ่วงน้ำหนักข้อมูลก่อน แล้วจึงนำข้อมูลทั้งหมดที่ถ่วงน้ำหนักแล้วมาจัดอันดับของข้อมูลในภายหลัง ซึ่งต่อมาสถิติทดสอบนี้จึงใช้ชื่อว่าสถิติทดสอบฮอดเจส-เลย์แมน อะโลนเม้นท์ (Hodges-Lehmann Alignment test)

สถิติทดสอบฮอดเจส-เลย์แมน อะโลนเม้นท์

เป็นสถิติทดสอบที่ได้รับการปรับปรุงมาจากรูปแบบของวิลค็อกซอน ซึ่งใช้ทดสอบสมมติฐาน เพื่อเปรียบเทียบอันดับของประชากรตั้งแต่ 2 กลุ่มขึ้นไป โดยจะต้องถ่วงน้ำหนักข้อมูลเพื่อจัด

ความแตกต่างระหว่างบล็อกเสียก่อน จากนั้นจึงนำข้อมูลที่ได้ทั้งหมดมาจัดเรียงอันดับจากน้อยไปมากแล้วทดสอบสมมติฐานต่อไป

จากโมเดลการวิเคราะห์ความแปรปรวนแบบกลุ่มสุ่ม เมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่าง ตัวแปรบล็อก (block) กับตัวแปรทดลอง (treatment)

$$X_{bki} = \mu + \alpha_k + \beta_{b..} + \epsilon_{bki} \quad \begin{array}{l} i = 1, 2, \dots, n \\ k = 1, 2, \dots, K \\ b = 1, 2, \dots, B \end{array}$$

ข้อตกลงเบื้องต้น

1. ข้อมูลสามารถจัดให้อยู่ในรูปของอันดับได้
2. ข้อมูลมีการแจกแจงความน่าจะเป็นแบบต่อเนื่อง
3. กลุ่มตัวอย่างต้องเป็นอิสระจากกัน
4. ไม่มีปฏิสัมพันธ์ระหว่างตัวแปรบล็อกกับตัวแปรทดลอง

สมมติฐาน : สมมติฐานที่ต้องการทดสอบ คือ $H_0 : E(\bar{R}_1) = E(\bar{R}_2) = E(\bar{R}_3) \dots$

วิธีการคำนวณ : ถ้ากำหนดให้ข้อมูลที่ถ่วงน้ำหนักแล้วเป็น X'_{bki}

ดังนั้น

$$X'_{bki} = X_{bki} - \beta_{b..} = \mu + \alpha_k + \epsilon_{bki}$$

โดยที่ $\beta_{b..}$ ประมาณได้จาก $\hat{\beta}_{b..} = (\bar{X}_{b..} - \bar{X}_{...})$

ดังนั้น X'_{bki} จะประมาณได้จาก $X'_{bki} = X_{bki} - \hat{\beta}_{b..} = X_{bki} - (\bar{X}_{b..} - \bar{X}_{...})$

ฉะนั้นข้อมูลที่ถ่วงน้ำหนักแล้วคือ $X'_{bki} = X_{bki} - \bar{X}_{b..} + \bar{X}_{...}$

เนื่องจากค่าเฉลี่ยรวม (grand means: $\bar{X}_{...}$) เป็นค่าคงที่ ดังนั้นข้อมูลที่ถ่วงน้ำหนักแล้ว

สามารถประมาณได้จาก $X'_{bki} = X_{bki} - \bar{X}_{b..}$

นำข้อมูลที่ถ่วงน้ำหนักแล้วทั้งหมดมาจัดเรียงอันดับจากน้อยไปมาก และ กำหนดให้

r_{bki} คืออันดับของข้อมูลที่ถ่วงน้ำหนักแล้ว

$S_{b..}$ คือความแปรปรวนของอันดับในบล็อก (variance of the ranks in block b)

ดังนั้น

$$\bar{R}_{...} = \frac{1}{nBK} \sum_{b=1}^B \sum_{k=1}^K \sum_{i=1}^n r_{bki}$$

$$\bar{R}_{b..} = \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n r_{bki}$$

$$\bar{R}_{.k} = \frac{1}{nB} \sum_{b=1}^B \sum_{i=1}^n r_{bki}$$

$$S_{b..}^2 = \frac{\sum_{k=1}^K \sum_{i=1}^n (r_{bki} - \bar{R}_{b..})^2}{nK}$$

$$\bar{S}_b^2 = \frac{1}{B} \sum_{b=1}^B S_{b..}^2$$

สถิติทดสอบ

$$W = \frac{nK - 1}{nK} \sum_{k=1}^K \left[\frac{\bar{R}_{.k} - \bar{R}_{...}}{\bar{S}_b / \sqrt{nB}} \right]^2$$

ซึ่งมีการแจกแจงแบบไคสแควร์ (chi-square) ที่มีชั้นแห่งความเป็นอิสระ $\nu = K - 1$.

จากการศึกษาของฮอดเจส และเลย์แมน (Hodges and Lehmann, 1962) พบว่า ประสิทธิภาพของสถิติทดสอบฮอดเจส-เลย์แมน อะโลนเม้นท์ เมื่อใช้วิธีการถ่วงน้ำหนักข้อมูลด้วย ค่าเฉลี่ยเลขคณิตของแต่ละบล็อก และจำนวนตัวแปรทดลองเท่ากับ 2 จะมีประสิทธิภาพไม่ต่ำกว่า 0.955 เมื่อเปรียบเทียบกับสถิติทดสอบเอฟ (F-test)

งานวิจัยที่เกี่ยวข้อง

Dixon (1954) ได้ศึกษาเปรียบเทียบ เค-เอส เทส (K-S test) กับแรงค์ ซัมเทส (rank sum test), มีเดียน เทส (median test), รันส์ เทส (runs test) และที เทส (t-test) แบบสองทาง ผลปรากฏว่าดีสตริบิวชันฟรี (distribution-free) ทั้ง 4 วิธี มีอำนาจ

ในการทดสอบสูงกว่า การทดสอบที่เมื่อกลุ่มตัวอย่างมีขนาดเล็ก ซึ่งในการศึกษาเปรียบเทียบนี้ใช้ขนาดของกลุ่มตัวอย่างเท่ากับ 5 มีลักษณะเป็นแบบปกติ ที่ระดับนัยสำคัญ 0.025 สรุปได้ว่าการทดสอบโดยใช้อันดับ (rank test) มีอำนาจของการทดสอบดีกว่า เค-เอส เทส ส่วนเค-เอส เทส จะมีอำนาจการทดสอบดีกว่ามีเดียนเทส

Hodges และ Lehmann (1956, cited by Blair & Higgins 1980:311) ได้ศึกษาจุดที่น่าสนใจและศักยภาพที่สำคัญของ asymptotic เขาได้พิสูจน์ว่า ค่า A.R.E. (หรือ Pitman efficiency) ของการทดสอบของวิลค็อกซอนซึ่งสัมพันธ์กับการทดสอบที่มีค่าสูงถึงอินฟินิตี้ และมีค่าไม่ต่ำกว่า 0.864 จากผลที่ได้นี้ Hodges และ Lehmann (1956 : 356) ได้กล่าวว่

จากการทดสอบที่แล้วมาทำให้ทราบถึงประสิทธิภาพและผลที่ได้จากการใช้กลุ่มตัวอย่างที่มีขนาดต่างๆ กัน และแนวทางที่จะเป็นไปได้ในทางปฏิบัติทั้งหมด แสดงให้เห็นว่า การที่จะใช้การทดสอบของวิลค็อกซอนแทนการทดสอบที่ จะไม่ทำให้เกิดการสูญเสียประสิทธิภาพต่ออย่างไร และ ในทางตรงกันข้ามการทดสอบของวิลค็อกซอนอาจจะมีประสิทธิภาพในการทดสอบสูงกว่าการทดสอบที่

Boneau (1962, cited by Blair & Higgins 1980 : 312) ได้ใช้เทคนิคมอนติคาร์โลศึกษาเปรียบเทียบอำนาจของการทดสอบที่กับการทดสอบของวิลค็อกซอน เมื่อกลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็น normal , rectangular และ exponential ได้ข้อสรุปว่าโดยทั่วไปแล้วการทดสอบที่มีอำนาจของการทดสอบสูงกว่าการทดสอบ Mann-Witney U test (Wilcoxon test) แต่ไม่มากนัก

Neave และ Granger (1968, cited by Blair & Higgins 1980 : 312) ได้ศึกษาเปรียบเทียบอำนาจของการทดสอบที่กับการทดสอบของวิลค็อกซอน เมื่อกลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะการแจกแจงเป็นแบบปกติ และกลุ่มตัวอย่างมีขนาด $n_1 = n_2 = 20$ และ $n_1 = 20, n_2 = 40$ ตามลำดับ ผลปรากฏว่า การทดสอบของวิลค็อกซอนเมื่ออำนาจของการทดสอบเหนือกว่าการทดสอบที่ โดยที่การทดสอบของวิลค็อกซอนมีส่วนของการเกิดนัยสำคัญมากกว่าการทดสอบที่ และผลต่างของสัดส่วนของการเกิดนัยสำคัญของสถิติทดสอบทั้งสอง สูงถึง 0.12

Derek Srisukho (1974 : 1-3) ใช้เทคนิคมอนติคาร์โล เพื่อศึกษาเปรียบเทียบอำนาจของการทดสอบเอฟกับเอช เทสของดราสค์ลวอลลิสจากข้อมูลที่มีการแจกแจงของประชากร 6 รูป แบบได้แก่ แบบปกติ, แบบยูนิฟอร์ม, แบบ Double Exponential, แบบ Moderate Negatively Skewed, แบบ Extreme Negatively Skewed และแบบ Bimodal ขนาดกลุ่มตัวอย่างเท่ากับ 5, 10, 15 และ 20 ทำการจำลองการทดลองซ้ำ 1,000 ครั้ง และตั้งระดับนัยสำคัญที่เท่ากับ .01 และ .05 ผลการศึกษาสรุปได้ว่า สำหรับการแจกแจงของประชากรที่มีลักษณะ leptokurtic เอช เทส จะมีอำนาจของการทดสอบสูงกว่าการทดสอบเอฟ แต่ถ้าการแจกแจงของประชากรมีลักษณะ platykurtic และแบบปกติแล้ว เอช-เทส จะมีอำนาจของการทดสอบต่ำกว่าการทดสอบเอฟ

Blair, Higgins และ Smitley (1980, cited by Blair & Higgins 1980 : 312) ได้ทำการซิมูเลท (Simulate) โดยใช้คอมพิวเตอร์ เพื่อเปรียบเทียบอำนาจของการทดสอบที่กับการทดสอบของวิลค็อกซอน เมื่อประชากรมีลักษณะการแจกแจงเป็นแบบ exponential เขาได้สรุปว่าการทดสอบของวิลค็อกซอนมีอำนาจของการทดสอบเหนือกว่าการทดสอบที่

Blair และ Higgins (1980 : 309-335) ได้ใช้เทคนิคมอนติคาร์โล ทำการศึกษา เพื่อเปรียบเทียบอำนาจของการทดสอบของวิลค็อกซอน กับการทดสอบที่ ภายใต้ลักษณะการแจกแจง ของประชากรแบบ Uniform ,แบบ laplace ,แบบ half normal , แบบ exponential , แบบ mixed normal และแบบ mixed-uniform โดยกลุ่มตัวอย่างที่ใช้ในการศึกษามีขนาด (n_1, n_2) เท่ากับ (3,9), (6,6), (9,27), (18,18), (27,81) และ (54,54) ผลการศึกษา สรุปได้ว่า

1. โดยทั่วไปการทดสอบของวิลค็อกซอนมีอำนาจของการทดสอบสูงกว่าการทดสอบที่มาก
2. A.R.E. เป็นดัชนีที่ดีในการชี้ถึงอำนาจของการทดสอบทั้งสอง
3. ผลการศึกษาจากกลุ่มตัวอย่างที่มีขนาดเล็ก จะได้ผลที่แตกต่างไปจากการศึกษาใน กรณีที่กลุ่มตัวอย่างมีขนาดใหญ่
4. เนื่องจากการศึกษาที่มีมาก่อนนี้ ศึกษาเกี่ยวกับประชากรที่มีการกระจายน้อยและใช้กลุ่ม ตัวอย่างที่แตกต่างกันออกไปมาก ทำให้ผลที่ได้มาอยู่ในภาวะที่น่าสงสัย

จากงานวิจัยของโบรมา (2526) ใช้เทคนิคมอนติคาร์โล ศึกษาเปรียบเทียบอำนาจ ของการทดสอบที, การทดสอบของวิลค็อกซอน, การทดสอบเทอซี-โฮพฟ์ดิง นอร์มอล-สกอรี และการทดสอบแวน-เดอแวนร์เดน นอร์มอล-สกอรี เมื่อกลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะ การแจกแจงเป็นแบบปกติ แบบยูนิฟอร์ม และแบบโลจิสติก ได้ข้อสรุปว่า การทดสอบที่มีอำนาจของ การทดสอบสูงที่สุดในการทดสอบทั้ง 4 วิธี เมื่อลักษณะข้อมูลที่ใช้ในการทดสอบอยู่ในมาตราอันตรภาค หรือมาตราอัตราส่วน สำหรับการทดสอบของเทอซี-โฮพฟ์ดิง นอร์มอล-สกอรีและการทดสอบ ของ แวนเดอ แวนร์เดน นอร์มอล-สกอรี มีอำนาจของการทดสอบเหนือกว่าการทดสอบของวิลค็อกซอน เมื่อลักษณะข้อมูลที่ใช้ในการทดสอบอยู่ในรูปของอันดับ