บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวิจัยเป็นวิถีชีวิตของมนุษย์ มนุษย์ใช้การวิจัยสำหรับแสวงหาความรู้มาเป็นเวลานาน โดยเฉพาะในปัจจุบันการวิจัยทางด้านต่างๆ มีบทบาทที่สำคัญต่อการดำเนินชีวิตของคนเราทั้งทางตรงและทางอ้อมเป็นอย่างมาก เนื่องจากการวิจัยก่อให้เกิดข้อความรู้ใหม่ที่เป็นคำตอบของปัญหาต่าง ๆได้ การวิจัยเกี่ยวกับปัญหาด้านต่าง ๆ มีวิธีดำเนินการได้หลายวิธีด้วยกัน แต่วิธีหนึ่งในจำนวนหลายวิธีที่มีการยอมรับและเชื่อถือกันมากในขณะนี้ คือ การวิจัยที่อาศัยข้อมูลและระเบียบวิธีทางสถิติเข้ามาช่วย (สรชัย พิศาลบุตร, 2528) ดังนั้น วิธีการดังกล่าวจึงถูกนำมาใช้อย่างกว้างขวางกับงานวิจัยทางสังคมศาสตร์และทางวิทยาศาสตร์ทุกสาขา ดังที่ นงเยาว์ เดชมา (2529,2) กล่าวว่า ระเบียบวิธีทางสถิติ เป็นวิธีการทางวิทยาศาสตร์ (Scientific Method) ที่ใช้เป็นเครื่องมือวินิจฉัยเกี่ยวกับความถูกต้องของสารสนเทศ เพื่อหาผลสรุปที่ถูกต้องของการศึกษาค้นคว้าสำหรับช่วยในการตัดสินใจอย่างมีประสิทธิภาพ ระเบียบวิธีการทางสถิติที่ใช้ในงานวิจัยต่าง ๆ มีขั้นตอนสำคัญสรุปได้ดังนี้

1. การเก็บรวบรวมข้อมูล (Data Collection) วิธีการเก็บรวบรวมข้อมูลทำได้หลายวิธี เช่น การเก็บรวบรวมข้อมูลจากเอกสารหลักฐาน การสังเกต การสัมภาษณ์ การใช้แบบสอบถามและแบบทดสอบ เป็นต้น ซึ่งแต่ละวิธีเหมาะสมกับลักษณะข้อมูล และแผนแบบการวิจัยแตกต่างกัน ผู้วิจัยควรเลือกใช้โดยคำนึงถึงวิธีการ จุดเด่นจุดด้อยของแต่ละวิธี รวมทั้งความจำกัดของทรัพยากร แรงงาน และเวลาของตนด้วย

2. การประมวลผลและการนำเสนอข้อมูล (Data Processing and Presentation) เมื่อได้ข้อมูลแล้ว สถิติจะมีบทบาทในการประมวลผล และนำเสนอข้อมูลที่เก็บรวบรวมมาได้ให้อยู่ในรูปแบบที่น่าสนใจ เข้าใจความหมาย และอธิบายปรากฏการณ์ที่ง่ายต่อการตีความหมาย และน่าติดตาม เช่น การสรุปตัวเลข การทำตาราง การทำแผนภาพ เป็นต้น

3. การวิเคราะห์ข้อมูล (Data Analysis) มีเทคนิคทางสถิติมากมายที่ใช้สำหรับวิเคราะห์ข้อมูล แต่การจะใช้เทคนิคใดในการวิเคราะห์ข้อมูลนั้นจะต้องพิจารณาที่วัตถุประสงค์ และความสอดคล้องระหว่างลักษณะของข้อมูลกับข้อตกลงเบื้องต้นของสถิติต่าง ๆ เหล่านี้

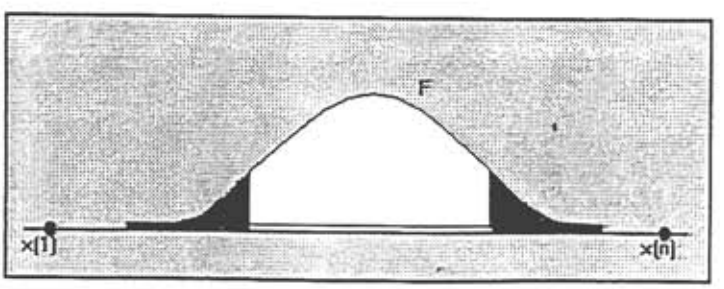
4. การตีความหมายข้อมูล (Data Interpretation) เป็นผลสืบเนื่องมาจากการวิเคราะห์ข้อมูล เนื่องจากข้อสรุปจากการวิเคราะห์ จะมีผลทางตัวเลขเท่านั้น จึงต้องอาศัยหลักการในเนื้อเรื่องที่ทำการวิจัย เป็นแนวทางในการอธิบาย เพื่อขยายความจากผลลัพธ์ทางตัวเลขให้ง่ายแก่การเข้าใจ

จากขั้นตอนของระเบียบวิธีทางสถิติดังกล่าวจะเห็นได้ว่า การเก็บรวบรวมข้อมูลเป็นขั้นตอนที่สำคัญ ผู้วิจัยจะต้องลงมือปฏิบัติ ภายหลังจากที่กำหนดปัญหาวิจัยและออกแบบการวิจัยไว้แล้ว ทั้งนี้เพราะว่า การเก็บรวบรวมข้อมูลเป็นขั้นตอนที่ทำให้ผู้วิจัยได้ข้อมูลสำหรับนำไปวิเคราะห์ และสรุปผลการวิจัย ถ้าข้อมูลที่ได้เป็นค่าสังเกตที่มีคุณภาพย่อมส่งผลให้การวิจัยมีคุณภาพ แต่ถ้าค่าสังเกตที่ได้เป็นค่าสังเกตที่คลาดเคลื่อนหรือผิดปกติก็จะทำให้ผลการวิจัยคลาดเคลื่อนจากความเป็นจริงไปด้วย ดังที่มีคำกล่าวว่า "Garbage in Garbage out"

ค่าสังเกตที่มีลักษณะผิดปกติ อาจจำแนกได้เป็น 4 แบบ คือ ค่าสังเกตแผกเหล่า (outlying observation) ค่าสังเกตสุดโต่ง (extreme observation) ค่าสังเกตปนเปื้อน (contaminate observation) และค่าสังเกตสูญหาย (missing observation)

1. ค่าสังเกตแผกเหล่า คือ ค่าสังเกตที่มีค่าไม่สอดคล้องกับค่าสังเกตอื่นที่เป็นไปได้ทั้งหมดของชุดข้อมูล เช่น ในการเก็บรวบรวมข้อมูลผู้วิจัยทำการสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงแบบ F ได้ค่าสังเกต $X_{(1)}$, $X_{(2)}$, ... , $X_{(n)}$ ตามภาพ

ภาพ ก.



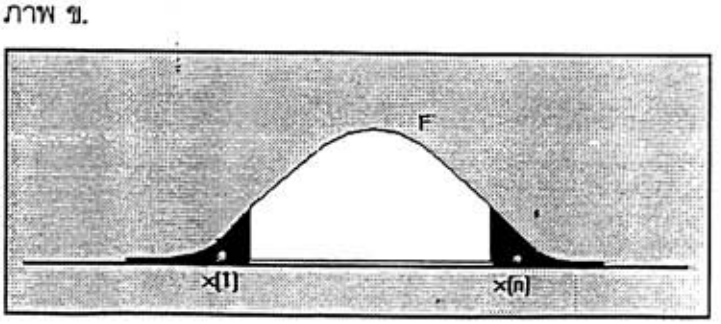
จากภาพ ก. จะเห็นได้ว่า $X_{(1)}$ และ $X_{(n)}$ เป็นค่าแผกเหล่า เนื่องจากเป็นค่าสังเกตที่อยู่ภายนอกหรือไม่สัมพันธ์กับลักษณะการแจกแจง F ดังกล่าว

นอกจากความหมายของค่าแผกเหล่าดังกล่าว Grubbs (1964) ได้ให้ความหมายของค่าแผกเหล่าไว้ว่า ค่าแผกเหล่า คือ ค่าสังเกตที่มีลักษณะเบี่ยงเบนไปจากค่าสังเกตที่เป็นไปได้ทั้งหมดในชุดข้อมูลนั้น และ Manrice and William (1960) และ Gentleman and Wilk (1975) กล่าวว่า ในกลุ่มตัวอย่างที่ประกอบด้วยค่าสังเกต n ค่า มีความเป็นไปได้ที่ผู้วิจัยจะกำหนดเกณฑ์สำหรับศึกษาความเบี่ยงเบนของค่าสังเกตใด ๆ ที่เบี่ยงเบนกว่าค่าที่เหลือทั้งหมดในชุดข้อมูล เนื่องจากค่าเหล่านี้มักจะก่อให้เกิดความสงสัยว่าเป็นค่าสังเกตที่มาจากประชากรอื่น หรือวิธีการเก็บรวบรวมข้อมูลมีความผิดพลาด ซึ่งค่าเหล่านี้จะถูกเรียกว่าค่าแผกเหล่า และสามารถใช้วิธีการทดสอบ (test) เพื่อค้นหาว่าค่าแผกเหล่ามีความเป็นเอกพันธ์ (homogeneous) กับค่าสังเกตอื่นในชุดข้อมูลหรือไม่

จากการศึกษาของ Barnett and Lewis (1983) ได้ให้ข้อมูลเสนอว่า ค่าแผกเหล่าจะทำให้ความคลาดเคลื่อนมาตราฐาน (S. E.) มีค่ามากขึ้น และ Daniel (1978) กล่าวว่าค่าแผกเหล่าไม่ได้ทำให้เกิดความลำเอียงเฉพาะค่าของตัวเอง และค่าส่วนที่เหลือเท่านั้น แต่จะมีอิทธิพลทั้งแถวและหลักที่มีค่าสังเกตแผกเหล่าปรากฏอยู่ และส่งผลให้ค่าสถิติที่ได้มีความถูกต้องลดลง

2. ค่าสังเกตสุดโต่ง คือ ค่าสังเกตที่มีค่ามากที่สุด หรือน้อยที่สุดของค่าที่เป็นไปได้ทั้งหมดในชุดข้อมูล หรือกล่าวได้ว่า ค่าสังเกตสุดโต่ง คือ ค่าสังเกตที่มีตำแหน่งอยู่ที่ส่วนปลายสุดของการแจกแจงข้อมูล ตามภาพ

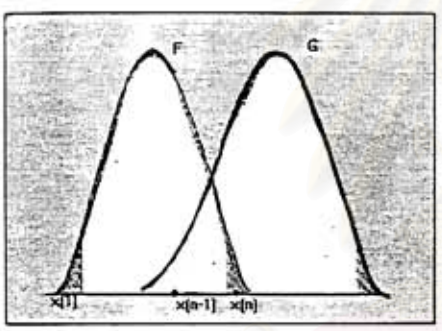
ภาพ ข.



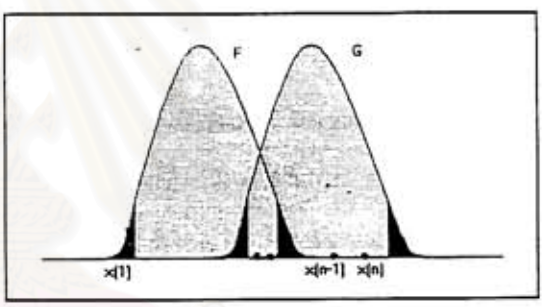
จากภาพ ข. อธิบายได้ว่า $X_{(1)}$ และ $X_{(n)}$ เป็นค่าสังเกตสุดโต่งทางด้านน้อย และ ทางด้านมาก ตามลำดับ หรือเมื่อพิจารณาข้อมูลที่มีลักษณะการแจกแจงแบบปกติ (Normal Distribution) ค่าสังเกตใด ๆ ที่ระดับความเชื่อมั่น 99.999 % สามารถอธิบายได้ว่า ค่าสังเกตเหล่านั้น ยังคงเป็นค่าสังเกตสุดโต่งไม่ใช่ค่าแผกเหล่า

ข้อมูลชุดใดที่มีค่าสังเกตสุดโต่ง รวมอยู่ด้วยย่อมส่งผลให้การประมาณค่าพารามิเตอร์ รวมถึงการคำนวณค่าสถิติต่าง ๆ ที่สนใจ คลาดเคลื่อนไปจากความเป็นจริง

3. ค่าสังเกตปนเปื้อน คือ ค่าสังเกตที่ไม่ใช่ค่าที่แท้จริงของชุดข้อมูลที่ศึกษา ซึ่งสามารถแสดงลักษณะของค่าสังเกตปนเปื้อนจากแผนภาพต่อไปนี้



ภาพ ค.



ภาพ ง.

ให้ G เป็นการแจกแจงชุดข้อมูลอื่น ที่สัมพันธ์กับการแจกแจง F ซึ่งเป็นการแจกแจงของข้อมูลชุดที่ศึกษา จากภาพ ค. จะเห็นได้ว่า ค่าสังเกตปนเปื้อน (แทนด้วยสัญลักษณ์ .) คือ $X_{(n-1)}$ และ $X_{(n)}$ โดย $X_{(n-1)}$ เป็นค่าปนเปื้อนที่เป็นค่ากลางของชุดข้อมูลที่ศึกษา ส่วน $X_{(n)}$ เป็นค่าปนเปื้อนที่เป็นค่าสุดโต่ง แต่ทั้งสองค่าไม่เป็นค่าสังเกตแผกเหล่า และเมื่อพิจารณาภาพ ง. จะพบว่า ทั้ง $X_{(n-1)}$ และ $X_{(n)}$ เป็นค่าปนเปื้อน ซึ่งจะเห็นได้ว่า $X_{(n-1)}$ ไม่ใช่ค่าสุดโต่งของชุดข้อมูลแต่ทั้ง $X_{(n-1)}$ และ $X_{(n)}$ เป็นค่าสังเกตแผกเหล่า

ค่าสังเกตปนเปื้อนที่รวมอยู่ในชุดข้อมูล อาจทำให้ลักษณะการแจกแจงของข้อมูลผิดปกติ ซึ่งจะส่งผลให้การทดสอบสมมติฐานและการคำนวณค่าสถิติต่าง ๆ คลาดเคลื่อนไปด้วย

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. ค่าสังเกตสูญหาย เกิดจากการที่ผู้วิจัยไม่สามารถรวบรวมข้อมูลเกี่ยวกับตัวแปรที่ศึกษาได้ครบถ้วน เมื่อเสร็จสิ้นงานวิจัย ซึ่งอาจเกิดขึ้นได้หลายกรณี เช่น พลวิจัยเกิดเจ็บป่วยหรือตาย ขณะทำการทดลอง หรือเกิดอพยพย้ายที่อยู่ หรือเกิดจากการไม่ตอบสนองของพลวิจัยต่อการวัด หรือการเก็บรวบรวมข้อมูล หรือเกิดจากการที่ผู้วิจัยไม่ได้บันทึกข้อมูลในวาระนั้น

จากการศึกษาของ Cohen and Cohen (1983) พบว่า ถ้าข้อมูลสูญหายมีลักษณะแบบสุ่ม การทดสอบความสัมพันธ์ของผลลัพธ์ที่ได้จะไม่ลำเอียงไปจากสมมติฐานเมื่อมีข้อมูลครบถ้วน แต่จะมีผลทำให้เกิดความคลาดเคลื่อนของชั้นความเป็นอิสระ เป็นผลให้อำนาจการทำสอบลดลง และยังมีผลให้ค่าความคลาดเคลื่อนมาตรฐานใหญ่ขึ้น

เมื่อเปรียบเทียบค่าสังเกตที่มีลักษณะผิดปกติทั้ง 4 แบบ จะเห็นว่าค่าแผกเหล่า ค่าสุดโต่ง และค่าปนเปื้อน มีลักษณะคาบเกี่ยวกันสรุปได้ 4 ลักษณะดังนี้ ค่าสุดโต่งไม่จำเป็นต้องเป็นค่าแผกเหล่า แต่ค่าแผกเหล่าทุกค่าเป็นค่าสุดโต่ง หรือสัมพันธ์กับค่าสุดโต่ง ค่าปนเปื้อนไม่จำเป็นต้องเป็นค่าสุดโต่งหรือค่าแผกเหล่า และค่าแผกเหล่าก็ไม่จำเป็นต้องเป็นค่าปนเปื้อน

เนื่องจากค่าสังเกตที่มีลักษณะผิดปกตินั้น เมื่อเกิดขึ้นในชุดของข้อมูลที่ศึกษาย่อมส่งผลต่อค่าการวิเคราะห์ข้อมูลดังที่กล่าวมาแล้ว แต่จากความหมายและลักษณะของค่าสังเกตผิดปกติทั้ง 4 แบบดังกล่าว จะเห็นว่า ค่าสูญหาย และค่าสุดโต่ง เป็นลักษณะผิดปกติที่ผู้วิจัยสามารถสังเกตและหาแนวทางตรวจสอบได้ และผู้วิจัยสามารถวางแนวทางปฏิบัติสำหรับกรณีที่เกิดข้อมูลลักษณะดังกล่าวขึ้น เช่น ใช้การประมาณค่าสังเกตสูญหายหรือใช้การประมาณค่าแทนข้อมูลสุดโต่งที่เกิดขึ้น สำหรับค่าปนเปื้อน Barnett and Lewis (1982) กล่าวว่าไม่มีทางที่จะทราบได้ว่าเกิดค่าปนเปื้อนขึ้นในชุดของข้อมูลที่เราศึกษา แต่สิ่งที่เราควรให้ความสนใจอย่างจริงจัง คือ ค่าแผกเหล่า เพราะเป็นไปได้ที่ค่านี้จะแสดงลักษณะที่ปนเปื้อนของชุดข้อมูล ดังนั้นวิธีการตรวจหาค่าแผกเหล่า โดยใช้วิธีการทางสถิติที่เหมาะสมจึงเป็นปัญหาที่ควรทำการศึกษาในปัจจุบัน

วิธีการตรวจหาค่าแผกเหล่า แยกตามประเภทของการวิจัย และจำนวนตัวแปรที่มีค่าสังเกตแผกเหล่าได้ 4 วิธี คือ

1. การตรวจหาค่าแผกเหล่าในการวิจัยสำรวจ เมื่อมีตัวแปรที่ศึกษาเพียง 1 ตัว
2. การตรวจหาค่าแผกเหล่าในการวิจัยสำรวจ เมื่อมีตัวแปรที่ศึกษาหลายตัว
3. การตรวจหาค่าแผกเหล่าในการวิจัยเชิงทดลอง
4. การตรวจหาค่าแผกเหล่าในการวิจัยแบบลำดับเวลา

การวิจัยทั้ง 3 ลักษณะ คือ การวิจัยสำรวจ การวิจัยทดลอง และการวิจัยแบบลำดับขั้นเวลา จะเห็นได้ว่า การวิจัยสำรวจ และวิจัยแบบลำดับขั้นเวลา เป็นการวิจัยที่มักทำการศึกษากับกลุ่มตัวอย่างขนาดใหญ่ ผลกระทบของค่าแผกเหล่าในการวิจัยสำรวจและการวิจัยแบบลำดับขั้นเวลา ย่อมมีน้อยกว่าในการวิจัยทดลอง ซึ่งมักจะศึกษากับกลุ่มตัวอย่างที่มีขนาดเล็ก และเนื่องจากการวิจัยเชิงทดลองผู้วิจัยมีความต้องการที่จะเปรียบเทียบความแตกต่างจากทรีตเมนต์ที่ใช้ในการทดลอง ถ้าค่าสังเกตที่ได้จากการเก็บรวบรวมข้อมูลเป็นค่าสังเกตแผกเหล่าเสียแล้ว ย่อมทำให้ผลที่ได้ผิดพลาด และเป็นอันตรายต่อการนำผลการวิจัยไปใช้ ดังนั้น การศึกษาการตรวจหาค่าสังเกตแผกเหล่าในการวิจัยเชิงทดลองจึงเป็นระเบียบวิธีที่สำคัญยิ่งสำหรับความตรงในการสรุปผลการวิจัยเชิงทดลอง

จากการรายงานเอกสารที่เกี่ยวข้องกับการตรวจหาค่าแผกเหล่าในแผนแบบวิจัยเชิงทดลอง พบว่า การตรวจหาจะใช้ค่าส่วนที่เหลือ (residual) เป็นเกณฑ์ในการพิจารณา โดย Daniel (1960) ได้ทำการศึกษาและเสนอแนวคิดว่าเซลล์ใดที่มีค่าส่วนที่เหลือสูง ๆ เซลล์นั้นจะเป็นค่าแผกเหล่า ต่อมาได้มีผู้พัฒนาวิธีการตรวจหาค่าแผกเหล่าในวิธีที่แตกต่างกันออกไป แต่ยังคงใช้พื้นฐานของค่าส่วนที่เหลือในการพิจารณา ซึ่งสามารถสรุปได้เป็น 3 วิธีที่สำคัญ คือ

1. การตรวจหาค่าแผกเหล่าโดยใช้ค่าสูงสุดของกำลังสองของค่าส่วนที่เหลือ
2. การตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ (Tetrads)
3. การตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ร่วมกับแผนภูมิเส้นกึ่งปกติ (Half - Normal Plot)

ทั้ง 3 วิธีเหมาะสมกับการวิจัยเชิงทดลอง ที่มีแผนแบบ 1 - k แฟคเตอร์ โดยที่แฟคเตอร์ทั้ง k ตัว ไม่มีปฏิสัมพันธ์ (interaction) ระหว่างกันหรือกล่าวอีกอย่างหนึ่งได้ว่าเหมาะสมกับแผนแบบการวิจัยเชิงทดลองแบบสุ่มสมบูรณ์ (Completely Randomized Design) แผนแบบการทดลองแบบบล็อคสมบูรณ์ (Randomized Complete Block Design) และแผนแบบการทดลองจตุรัสละติน (Latin Square Design) ได้มีผู้ทำการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจหาค่าสังเกตแผกเหล่าของทั้ง 3 วิธี ดังกล่าว และเสนอข้อความรู้ที่สำคัญไว้ ดังนี้

John และ Prescott (1975) และ Daniel (1978) ได้ทำการศึกษาและเสนอว่าวิธีการตรวจหาโดยใช้ค่าสูงสุดของกำลังสองค่าส่วนที่เหลือ จะสามารถตรวจหาค่าแผกเหล่าได้ถูกต้องเพียงค่าเดียว

Bradu และ Hawkins (1982) ได้ทำการศึกษาโดยใช้ค่าชุดสี่ในการตรวจหาค่าแผกเหล่า พบว่า การตรวจหาโดยใช้ค่าชุดสี่ สามารถตรวจหาค่าแผกเหล่าได้หลายค่า แต่ความถูกต้องชัดเจนยังมีน้อย ส่วนการตรวจหาโดยใช้ค่าชุดสี่ร่วมกับแผนภูมิเส้นกึ่งปกติ จะสามารถตรวจหาค่าแผกเหล่าได้ทั้งกรณีค่าแผกเหล่าเพียงค่าเดียว และหลายค่าได้อย่างถูกต้องเพียงพอ ซึ่งค่าชุดสี่นี้นอกจากจะสามารถตรวจสอบค่าแผกเหล่าได้หลาย ๆ ค่าแล้ว ยังสามารถใช้ประมาณค่าของค่าแผกเหล่า และค่าสังเกตสูญหาย (missing observation) โดยใช้ค่ามัธยฐานของค่าชุดสี่ได้อีกด้วย

วิธีการตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ (Tetrads) มี 4 ขั้นตอนที่สำคัญ คือ

ขั้นที่ 1 คำนวณค่าชุดสี่ จากสูตร $T_{ij.\alpha\beta} = Y_{ij} - Y_{\alpha j} - Y_{i\beta} + Y_{\alpha\beta}$

ขั้นที่ 2 หาค่ามัธยฐานของค่าชุดสี่ $(Q_{2(.,.)})$

ขั้นที่ 3 จัดลำดับค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่ $(|Q_{2(.,.)}|)$ จากมากไปน้อย

ขั้นที่ 4 พิจารณาเซลล์ที่มีค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่มาก ๆ เซลล์นั้นจะเป็นค่าแผกเหล่า

จากขั้นตอนการตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ จะเห็นได้ว่าวิธีการดังกล่าวยังไม่มีเกณฑ์ที่เหมาะสม ที่จะทำการวินิจฉัยว่า ขนาดของค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่ควรมีขนาดเท่าใดจึงจะถือว่าเซลล์นั้นเป็นค่าแผกเหล่า สำหรับวิธีการตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ร่วมกับแผนภูมิเส้นกึ่งปกติ ถึงแม้จะสามารถตัดสินจำนวนค่าแผกเหล่าได้ชัดเจนกว่าการใช้ค่าชุดสี่เพียงอย่างเดียว แต่วิธีการดังกล่าวนี้จะต้องนำค่ามัธยฐานของค่าชุดสี่ของทุกเซลล์มาคำนวณค่าความน่าจะเป็น และค่าความน่าจะเป็นสะสม เพื่อนำมาสร้างแผนภูมิเส้นกึ่งปกติระหว่างค่าความน่าจะเป็นสะสม กับค่าสัมบูรณ์ของค่ามัธยฐานชุดสี่ ซึ่งวิธีการดังกล่าวยังคงเป็นวิธีการที่ค่อนข้างยุ่งยาก สำหรับผู้วิจัยที่ต้องการตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ ดังนั้นจึงควรได้มีการทำการศึกษาเพื่อสร้างเกณฑ์ที่มีความสะดวกและเหมาะสม สำหรับใช้ตัดสินจำนวนค่าแผกเหล่าจากค่าชุดสี่ที่คำนวณได้อย่างถูกต้อง

ด้วยเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาวิธีการตรวจหาค่าแผกเหล่าในชุดข้อมูลโดยใช้ค่าชุดสี่ และทำการพัฒนาเกณฑ์สำหรับตัดสินจำนวนค่าแผกเหล่าที่ได้จากการคำนวณค่าชุดสี่ ซึ่งวิธีการเพื่อให้ได้ข้อสรุปดังกล่าวสามารถศึกษาได้โดยใช้การจำลองสถานการณ์ (Simulation) ด้วยเครื่องคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo) ซึ่งเป็นวิธีที่สามารถนำมาดำเนินการ เพื่อหาข้อสรุปในการทดลองได้ในปัจจุบัน

วัตถุประสงค์ของการวิจัย

1. เพื่อตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ (Tetrads) ในแผนการวิจัยเชิงทดลองแบบสุ่มสมบูรณ์
2. เพื่อเปรียบเทียบจำนวนค่าแผกเหล่าที่สามารถตรวจหาได้ถูกต้องจากเกณฑ์ที่พัฒนาขึ้น 5 เกณฑ์ คือ

2.1 $|DC\ 1|$ เมื่อ DC 1 คือค่ามัธยฐานชุดสี่ที่ระดับความเชื่อมั่น 99.00 % หรือ $z = 2.33$

2.2 $|DC\ 2|$ เมื่อ DC 2 คือค่ามัธยฐานชุดสี่ที่ระดับความเชื่อมั่น 99.50 % หรือ $z = 2.58$

2.3 $|DC\ 3|$ เมื่อ DC 3 คือค่ามัธยฐานชุดสี่ที่ระดับความเชื่อมั่น 99.90 % หรือ z = 3.09

2.4 $|DC\ 4|$ เมื่อ DC 4 คือค่ามัธยฐานชุดสี่ที่ระดับความเชื่อมั่น 99.99 % หรือ z = 3.72

2.5 $|DC\ 5|$ เมื่อ DC 5 คือค่ามัธยฐานชุดสี่ที่ระดับความเชื่อมั่น 99.999 % หรือ z = 4.27

3. เพื่อศึกษาลักษณะการแจกแจงของค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่

ขอบเขตการวิจัย

1. การวิจัยครั้งนี้มีค่าของตัวแปรดังต่อไปนี้

1.1 ตัวแปรตาม (Dependent Variable) คือ จำนวนค่าแผกเหล่าที่ตรวจหาได้จากเกณฑ์ $|DC\ 1|$, $|DC\ 2|$, $|DC\ 3|$, $|DC\ 4|$ และ $|DC\ 5|$ ที่ผู้วิจัยพัฒนาขึ้น

1.2 ตัวแปรต้น (Independent Variable) คือ วิธีตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ และเกณฑ์ $|DC\ 1|$, $|DC\ 2|$, $|DC\ 3|$, $|DC\ 4|$ และ $|DC\ 5|$ ซึ่งใช้สำหรับตัดสินจำนวนค่าแผกเหล่าที่ผู้วิจัยพัฒนาขึ้น

2. ศึกษาจากประชากรที่มีลักษณะการแจกแจงแบบปกติ (Normal Distribution) โดยจำกัดการศึกษาเฉพาะการแจกแจงของประชากรแบบปกติ ทางด้านขวา เท่านั้น

3. ศึกษาเฉพาะแผนแบบการวิจัยเชิงทดลองแบบสุ่มตลอด เมื่อ

3.1 กลุ่มตัวอย่างเป็น 3 , 5 และ 7 กลุ่ม

3.2 มีขนาดกลุ่มตัวอย่างเท่ากันทุกกลุ่ม คือ ขนาด 11 , 21 และ 31

4. กำหนดให้มีค่าแผกเหล่าแตกต่างกัน 3 ระดับ ดังนี้

4.1 ค่าแผกเหล่าระดับน้อย (x_1) เมื่อ

$$\bar{x} + 4.27 \text{ S.D.} < X_1 < \bar{x} + 4.99 \text{ S.D.}$$

4.2 ค่าแผกเหล่าระดับปานกลาง (x_2) เมื่อ

$$\bar{x} + 5.00 \text{ S.D.} < X_2 < \bar{x} + 5.99 \text{ S.D.}$$

4.3 ค่าแผกเหล่าระดับสูงมาก (x_3) เมื่อ

$$\bar{x} + 6.00 \text{ S.D.} < X_3 < \bar{x} + 6.99 \text{ S.D.}$$

โดยกำหนดให้มีจำนวนค่าแผกเหล่าเท่ากันทุกระดับ คือ 2 , 3 และ 4 ค่า

5. กำหนดให้มีเกณฑ์มาตราฐาน (Detective Criterion) สำหรับตัดสินค่าแผกเหล่าที่ได้จากการคำนวณค่าชุดสี่ 5 เกณฑ์ คือ

$|DC\ 1|$ = ค่าสัมบูรณ์ของค่ามัธยฐานของชุดสี่ที่ระดับความเชื่อมั่น 99.00 % ($z = 2.33$)

$|DC\ 2|$ = ค่าสัมบูรณ์ของค่ามัธยฐานของชุดสี่ที่ระดับความเชื่อมั่น 99.50 % ($z = 2.58$)

$|DC\ 3|$ = ค่าสัมบูรณ์ของค่ามัธยฐานของชุดสี่ที่ระดับความเชื่อมั่น 99.90 % ($z = 3.09$)

$|DC\ 4|$ = ค่าสัมบูรณ์ของค่ามัธยฐานของชุดสี่ที่ระดับความเชื่อมั่น 99.99 % ($z = 3.72$)

$|DC\ 5|$ = ค่าสัมบูรณ์ของค่ามัธยฐานของชุดสี่ที่ระดับความเชื่อมั่น 99.999 % ($z = 4.27$)

6. กำหนดค่าพารามิเตอร์ $\mu = 500$ และ $\sigma^2 = 100$

7. แต่ละกรณีของการวิจัยครั้งนี้ จะทำการทดลองซ้ำจำนวน 2,000 ครั้ง โดยใช้มอนตคาร์โลซมูเลชั่น (Monte Carlo Simulation Technique) และโปรแกรมย่อยสับรูทีน

คำจำกัดความ

ค่าแผกเหล่า (outlier) หมายถึง ค่าสังเกตที่มีค่าไม่สอดคล้องกับค่าสังเกตอื่นที่เป็นไปได้ทั้งหมดของชุดข้อมูล ซึ่งในการวิจัยครั้งนี้จำลองสถานการณ์ให้เกิดค่าสังเกตแผกเหล่า 3 ระดับ ได้แก่

1. ค่าสังเกตแผกเหล่าระดับน้อย คือ ค่าสังเกตที่มีค่าความเบี่ยงเบนอยู่ระหว่าง 4.27 ถึง 4.99 ของส่วนเบี่ยงเบนมาตรฐานของชุดข้อมูล คำนวณได้จาก

$$\bar{x} + 4.27 \text{ S.D.} < x < \bar{x} + 4.99 \text{ S.D.}$$

2. ค่าแผกเหล่าระดับปานกลาง คือ ค่าสังเกตที่มีค่าความเบี่ยงเบนอยู่ระหว่าง 5.00 ถึง 5.99 ของส่วนเบี่ยงเบนมาตรฐานของชุดข้อมูล คำนวณได้จาก

$$\bar{x} + 5.00 \text{ S.D.} < x < \bar{x} + 5.99 \text{ S.D.}$$

3. ค่าแผกเหล่าระดับมาก คือ ค่าสังเกตที่มีค่าความเบี่ยงเบนอยู่ระหว่าง 6.00 ถึง 6.99 ของส่วนเบี่ยงเบนมาตรฐานของชุดข้อมูล คำนวณได้จาก

$$\bar{x} + 6.00 \text{ S.D.} < x < \bar{x} + 6.99 \text{ S.D.}$$

การตรวจหาค่าแผกเหล่าโดยใช้ค่าชุดสี่ หมายถึง วิธีการทางสถิติที่ใช้ตรวจหาค่าแผกเหล่า โดยมีขั้นตอนคือ

1. คำนวณค่าชุดสี่ ($T_{ij.\infty}$) ในแต่ละเซลล์ของตาราง จากสูตร

$$T_{ij.\infty} = Y_{ij} - Y_{\infty j} - Y_{i\infty} + Y_{\infty\infty}$$

2. หาค่ามัธยฐานของค่าชุดสี่ ($Q_{2(j)}$)

3. หาค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่ ($|Q_{2(j)}|$)

4. จัดลำดับค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่จากมากไปน้อย และพิจารณาค่าสังเกตแผกเหล่า

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ได้ข้อความรู้เกี่ยวกับวิธีการตรวจหาค่าแผกเหล่าหลาย ๆ ค่าโดยใช้ค่าชุดสี่ (tetrads) ในแผนการวิจัยเชิงทดลองแบบสุ่มสมบูรณ์

2. สามารถหาเกณฑ์ที่มีความสะดวกและเหมาะสม เพื่อตัดสินจำนวนค่าแผกเหล่าที่ได้จากการคำนวณค่าชุดสี่

3. ทำให้ทราบลักษณะการแจกแจงค่าสัมบูรณ์ของค่ามัธยฐานของค่าชุดสี่

4. ผลพลอยได้จากการวิจัยครั้งนี้สามารถใช้เป็นแนวทางสำหรับศึกษาเกี่ยวกับวิธีตรวจหาค่าแผกเหล่าในแผนแบบการทดลองอื่น ตลอดจนใช้เป็นแนวทางในการศึกษาวิธีการพัฒนาเกณฑ์เพื่อตัดสินจำนวนค่าแผกเหล่า หรือทำการตรวจหาค่าแผกเหล่าโดยใช้วิธีอื่น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย