

บทที่ 1

บทนำ



ความเป็นมาและความสำคัญของปัญหา

ในการวิจัยเพื่อหาผลสรุปโดยทั่วไปที่ต้องทำการทดสอบสมมติฐานด้วยวิธีทดสอบทางสถิติ ส่วนใหญ่จะขึ้นอยู่กับการศึกษาจิตใจของผู้วิจัยในการที่จะเลือกใช้สถิติให้เหมาะสมกับข้อมูลที่จะทำการวิเคราะห์ ซึ่งการเลือกใช้สถิติแต่ละวิธีก็มักจะมีข้อตกลงเฉพาะของวิธีนั้นๆ เพื่อให้ นักวิจัยได้ตัดสินใจว่าเหมาะสมกับข้อมูลที่จะนำมาวิเคราะห์หรือไม่ โดยเฉพาะอย่างยิ่งในสถานการณ์ของการวิจัยเชิงทดลองที่นอกจากผู้วิจัยจะต้องมีความสามารถในการออกแบบการทดลองให้สามารถควบคุมตัวแปรในการทดลองได้แล้วสิ่งสำคัญและจำเป็นอีกประการหนึ่งของผู้วิจัยก็คือการวางแผนการวิเคราะห์ด้วยการเลือกใช้วิธีทดสอบทางสถิติที่เหมาะสมอันจะทำให้ได้ผลสรุปที่มีความคลาดเคลื่อนต่อความเป็นจริงเท่ากับโอกาสที่ผู้วิจัยยอมให้เกิดขึ้นในการทดสอบสมมติฐาน

การวิจัยเชิงทดลองจำนวนมาก มักจะเกี่ยวข้องกับ การให้การรักษา (Treatments) หลายๆวิธีและการสรุปผลเพื่อที่จะหาข้อสรุปว่า การรักษาใดมีประสิทธิภาพมากที่สุด ซึ่งตามปกติผู้วิจัยจะสรุปผลการวิจัยด้วยการทดสอบสมมติฐานและเทคนิคของการทดสอบสมมติฐานที่นำมาใช้กันมาก คือ การทดสอบเอฟหรือการวิเคราะห์ความแปรปรวน (ANOVA) ผลสรุปที่ได้จากการวิเคราะห์จะชี้ให้เห็นว่าในการทดลองกับกลุ่มตัวอย่างที่ให้การรักษาไม่เหมือนกันนั้นมีความแตกต่างกันบ้างหรือไม่ แต่ไม่ได้บ่งชี้ถึงรายละเอียดว่า ความแตกต่างที่เกิดขึ้นนี้เป็นผลมาจากการรักษาใด ซึ่งถ้าหากว่าผู้วิจัยต้องการที่จะตรวจสอบให้แน่ชัดว่าสาเหตุของความแตกต่างที่เกิดขึ้นนี้เป็นผลมาจากการรักษาใดแล้ว ก็จำเป็นต้องอาศัยกระบวนการวิเคราะห์ที่เพิ่มเติม ซึ่งโดยมากจะเกี่ยวข้องกับวิธีการเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างที่เรียกว่า การเปรียบเทียบพหุคูณ (Multiple Comparisons) ใน 2 ลักษณะคือ

1. การเปรียบเทียบภายหลัง (Posteriori Comparisons or Post hoc Analysis)
2. การเปรียบเทียบก่อนการวิเคราะห์ (Priori Comparisons or Planned Comparisons)

การเปรียบเทียบภายในแรก คือ กระบวนการเปรียบเทียบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่าง ซึ่งผู้วิจัยตั้งสมมติฐานในการวิเคราะห์ผลทดลองที่คาดว่าจะเกิดขึ้นก่อนที่จะทำการทดลอง ดังนั้นการวิเคราะห์ด้วยวิธีนี้จึงไม่จำเป็นที่ผู้วิจัยจะต้องทำการทดสอบเอฟก่อน เพราะเป็นความต้องการของผู้วิจัยที่จะสรุปผลที่บ่งบอกโดยเฉพาะที่ได้ตั้งไว้ในสมมติฐานก่อนเท่านั้น วิธีการของการเปรียบเทียบภายในแรกมีหลายวิธี เช่น การเปรียบเทียบภายในแรกอิสระด้วยอัตราส่วนที่ (Priori Orthogonal t Ratio Test) สถิติบอนเฟอโรนิตี (Bonferroni t Statistics) การเปรียบเทียบพหุคูณของตันน์ (Dunn's multiple Comparisons)

การเปรียบเทียบภายในหลัง (Posteriori Comparisons or Post hoc Analysis) คือการวิเคราะห์ข้อมูลภายในหลังจากการทดสอบเอฟแล้ว เพื่อหาผลสรุปว่าการเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างที่ให้ทริทเมนต์ต่างกันนั้นเป็นผลมาจากทริทเมนต์ใด และทริทเมนต์ใดบ้างที่แตกต่างกับทริทเมนต์อื่นอย่างมีนัยสำคัญ ดังนั้นการวิเคราะห์การเปรียบเทียบนี้จึงขึ้นอยู่กับวิธีการทดสอบเอฟ ถ้าการทดสอบเอฟไม่แสดงนัยสำคัญแล้ว ก็ไม่จำเป็นต้องวิเคราะห์การเปรียบเทียบภายในหลัง แต่ในกรณีที่การทดสอบเอฟมีนัยสำคัญ ก็จำเป็นจะต้องใช้วิธีการวิเคราะห์การเปรียบเทียบภายในหลัง ซึ่งมีหลายวิธี คือ LSD ของฟิชเชอร์, HSD ของทูกี, S* method ของเชฟเฟย์ วิธีการทดสอบช่วงพหุคูณของตันแคน (Duncan's new(multiple range Test) วิธีการทดสอบของตันเนตต์ (Dunnett's Test) และวิธีการทดสอบของนิวแมนคูลส์ (Newman Keuls Test) มโนทัศน์เบื้องต้น

ในการออกแบบการทดลองแบบกลุ่มสมบูรณ์ (CR-k) การทดสอบเอฟจะขึ้นอยู่กับโมเดล

$$x_{ij} = \mu + \beta_j + \epsilon_{1(j)}$$

ซึ่งมีความหมายว่า x_{ij} คือคะแนนของการสังเกตในการทดลองจะเกิดจากค่าเฉลี่ยประชากร (μ) บวกกับผลการทดลอง (β_j) และความคลาดเคลื่อนจากการทดลอง ($\epsilon_{1(j)}$) ซึ่งเป็นอิสระและข้อตกลงเบื้องต้นของการทดสอบเอฟมีดังต่อไปนี้

1. ประชากรมีการแจกแจงแบบปกติ
2. ความแปรปรวนของประชากรเท่ากัน
3. ประชากรแต่ละกลุ่มมีความเป็นอิสระต่อกัน

ผลของการฝ่าฝืนข้อตกลงเบื้องต้น

Cochran และ Cox (1957)¹ กล่าวว่า การขาดคุณสมบัติของข้อตกลงเบื้องต้นจะมีผลต่อระดับนัยสำคัญของการทดสอบและความไวของการทดสอบ ดังการทดสอบที่ตั้งระดับนัยสำคัญที่ .05 แต่ในความจริงอาจจะพบว่าอยู่ในระดับ .04 หรือ .07 การทดสอบที่มีความไวต่อการฝ่าฝืนข้อตกลงเบื้องต้น เช่นนี้ย่อมทำให้อำนาจของการทดสอบลดลงและผลสรุปที่ได้ อาจเกิดความคลาดเคลื่อนต่อความจริง

Scheff'e² กล่าวว่า ผลของการฝ่าฝืนข้อตกลงเบื้องต้นของความเป็นปกติมีผลเพียงเล็กน้อยต่อการอ้างอิงด้วยค่าเฉลี่ย แต่มีผลมากต่อการอ้างอิงด้วยความแปรปรวน และผลของการฝ่าฝืนข้อตกลงเบื้องต้น ความเท่ากันของความแปรปรวนจะมีผลเล็กน้อยต่อการอ้างอิงด้วยค่าเฉลี่ย ถ้าขนาดกลุ่มตัวอย่างเท่ากัน

จากการวิจัยและศึกษาของ Boneau (1960) Cochran (1947) Lindquist (1953) และ Bradley³ (1978) ที่ได้ทำการวิจัยเพื่อตรวจสอบความแกร่ง (robust) ของการทดสอบทางสถิติที่ใช้ในการวิเคราะห์ความแปรปรวนด้วยการทดสอบเอฟ ภายใต้สภาวะการของการกระจายไม่เป็นปกติ (nonnormality) และความไม่เท่ากันของความแปรปรวนของประชากร (heterogeneity of variance) การวิจัยดังกล่าวได้สรุปผลที่สอดคล้องกัน คือ การทดสอบเอฟมีลักษณะแกร่ง แม้ว่า การกระจายและความแปรปรวนของประชากรจะไม่เป็นไปตามข้อตกลงเบื้องต้นของการทดสอบเอฟ¹ ถึงแม้ว่าผลของการฝ่าฝืนข้อตกลงเบื้องต้นไม่ทำให้การทดสอบเอฟขาดความแกร่ง ดังนั้นการฝ่าฝืนข้อตกลงเบื้องต้นอาจจะมีผลต่อวิธีการเปรียบเทียบพหุคูณ

¹Roger E. Kirk, Experimental Design Procedures for the Behavioral Sciences, 2 nd ed. (California : Brooks Care Comp., 1969) p.78.

²Henry Scheffe, The Analysis of Variances, 6 ed. (New York : John Wiley 8 Sons, Inc., 1959), pp. 337-345.

³Phillip H. Ramsey, "Exact Type I Error Rates for Robustnes of Student's t Test with Unequal Variances," Journal of Educational Statistics, 5 (Winter 1980), pp. 337-349.

การควบคุมอัตราความคลาดเคลื่อน

ในการทดสอบสมมติฐานผู้ทดลองมีโอกาสที่จะเกิดความผิดจากการตัดสินใจผิด หรือที่เรียกว่าความคลาดเคลื่อนกล่าวคือในการที่ผู้ทดลองตัดสินใจปฏิเสธสมมติฐานที่ถูก ซึ่งในทางสถิติจะเรียกความผิดประเภทนี้ว่า ความคลาดเคลื่อนประเภทที่ 1 และเรียกความผิดที่เกิดจากการยอมรับสมมติฐานที่ผิดว่าความคลาดเคลื่อนประเภทที่ 2 ซึ่งในการทดสอบสมมติฐานผู้ทดลองจะพยายามป้องกันการตัดสินใจผิดพลาดทั้งสองประเภทสำหรับโอกาสที่จะยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 นั้น ผู้วิจัยสามารถระบุโอกาสเสี่ยงต่อการผิดพลาดไว้เท่ากับ α ซึ่งงานวิจัยทางด้านพฤติกรรมศาสตร์มักนิยมระบุ α ไว้เท่ากับ .05 และ .01

โดยปกติการทดลองที่มีเพียง 2 ทรีทเมนต์ และการทดสอบสมมติฐานที่มีการเปรียบเทียบเดียว ความคลาดเคลื่อนประเภทที่ 1 นี้จะถูกควบคุมด้วยการตั้งระดับนัยสำคัญของการทดสอบที่เรียกว่าระดับนัยสำคัญ α แต่สำหรับการเปรียบเทียบที่มีหลายทรีทเมนต์ หรือการทดสอบสมมติฐานที่มีการเปรียบเทียบมากกว่า 1 เช่นการเปรียบเทียบพหุคูณด้วยอัตราส่วนคู่ Ryan (1959) ได้กล่าวว่า ถ้าการทดลอง k ระดับ และทำการเปรียบเทียบแบบอิสระ $k-1$ จำนวนแล้วโอกาสที่จะเกิดการเปรียบเทียบที่มีนัยสำคัญผิดจะเท่ากับ $1-(1-\alpha)^C$ เมื่อ C หมายถึงจำนวนการเปรียบเทียบเช่นเมื่อทำการเปรียบเทียบค่าเฉลี่ย 3 กลุ่มและใช้การทดสอบที่กำหนดระดับนัยสำคัญ α เท่ากับ .05 ผลปรากฏว่า $1-(1-\alpha)^C = 0.975$ เท่ากับ 0.1426 สำหรับการเปรียบเทียบค่าเฉลี่ย 4 กลุ่ม... เท่ากับ 0.3017 สำหรับการเปรียบเทียบค่าเฉลี่ย 8 กลุ่ม ดังนั้น โอกาสที่จะเกิดอัตราความคลาดเคลื่อนนี้จะเพิ่มขึ้นตามจำนวนการเปรียบเทียบ ซึ่งแสดงให้เห็นว่าการควบคุมอัตราความคลาดเคลื่อนด้วยการแปลความหมายจากระดับนัยสำคัญเป็นสิ่งที่สับสนสำหรับอัตราความคลาดเคลื่อนของการทดลองที่มีหลายระดับ หรือการเปรียบเทียบที่มีหลายทรีทเมนต์ ดังนั้นจึงต้องมีหน่วยมโนทัศน์เกี่ยวกับอัตราความคลาดเคลื่อนโดยเฉพาะ เช่น อัตราความคลาดเคลื่อนต่อการเปรียบเทียบ (error rate per comparison) อัตราความคลาดเคลื่อนต่อสมมติฐาน (error rate per hypothesis) อัตราความคลาดเคลื่อนต่อการทดลองและชุดการทดลอง (error rate per experiment and experimentwise) อัตราความคลาดเคลื่อนต่อครอบครัวและต่อชุดของครอบครัว (error rate per family and familywise)

รายละเอียดและการคำนวณอัตราความคลาดเคลื่อนเหล่านี้ได้จาก Roger E. Kirk¹ *Experimental Design Procedures for the Behavioral Sciences*

เมื่อ τ หมายถึง อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง การวิจัยที่ศึกษาเกี่ยวกับอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองส่วนใหญ่จะตั้งเกณฑ์การตัดสินอัตราความคลาดเคลื่อน เพื่อที่จะบอกได้ว่าวิธีนั้นสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองได้เท่ากับอัตราความคลาดเคลื่อนที่ระบุหรือไม่ Cochran² (1954) ได้วิจัยเกี่ยวกับอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองของวิธีทดสอบไควสแควร์ และได้ตั้งเกณฑ์ในการตัดสินอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองด้วยลิมิต $.04 \leq \tau \leq .06$ สำหรับอัตราความคลาดเคลื่อนที่ระบุในระดับ $.05$ Bradley³ (1978) ได้แนะนำเกณฑ์ของการตัดสินความแกร่งด้วยลิมิต 2 แบบ คือ $0.9 \alpha \leq \tau \leq 1.1 \alpha$ ดังนั้นความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองจะอยู่ในช่วง $.045 \leq \tau \leq .055$ สำหรับอัตราความคลาดเคลื่อนที่ระบุในระดับ $.05$ และ $.009 \leq \tau \leq .011$ สำหรับอัตราความคลาดเคลื่อนที่ระบุในระดับ $.01$ และเกณฑ์ที่เรียกว่า Liberal Criteria ที่กำหนดอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองด้วยลิมิต $0.5 \alpha \leq \tau \leq 1.5 \alpha$ ดังนั้นลิมิต $.025 \leq \tau \leq .075$ สำหรับอัตราความคลาดเคลื่อนที่ระบุในระดับ $.05$ และลิมิต $.005 \leq \tau \leq .015$ สำหรับอัตราความคลาดเคลื่อนที่ระบุในระดับ $.01$ และลิมิต $.007 \leq \tau \leq .015$ สำหรับอัตราความคลาดเคลื่อนที่ระบุในระดับ $.01$

¹Kirk, *Experiment Design Procedures for the Behavioral Sciences*, 77.

²Phillip H. Ramsey, "Exact Type I Error Rates for Robustness of Student's t test with Unequal Variances," Journal of Educational Statistics, 5 NO. 4 (Winter 1980), pp. 337-349.

³Ibid.

สภาพปัจจุบันของปัญหา

เมื่อปัญหาของการฝ่าฝืนข้อตกลงเบื้องต้นซึ่งในส่วนของความเท่ากันของความแปรปรวนของประชากร อาจจะทดสอบได้จากสถิติ F_{max} แต่การฝ่าฝืนข้อตกลงของความเป็นปกติของการแจกแจงของประชากร ซึ่งผู้วิจัยไม่อาจจะทราบได้แน่ชัดถึงลักษณะการแจกแจงของประชากรที่ศึกษาว่าเป็นการแจกแจงในรูปใด และในการทดสอบทางสถิติเกี่ยวกับการทดสอบสมมติฐาน เช่นการทดสอบ ที่ การทดสอบไควสแควร์และการวิเคราะห์ความแปรปรวนด้วยการทดสอบเอฟ ซึ่งใช้ข้อตกลงของการทดสอบที่กล่าวถึงประชากรที่มีการแจกแจงแบบปกติ และความเท่ากันของความแปรปรวนประชากร แต่ในสถานการณ์ของความไม่จริงการแจกแจงของประชากรอาจแตกต่างจากข้อตกลงของการแจกแจงประชากรแบบปกติในลักษณะของความเบ้และความโด่ง แม้ว่าผู้วิจัยพยายามจะแก้ปัญหานี้โดยอาศัยทฤษฎี Central Limit Theorem โดยกำหนดให้ได้ใช้กลุ่มตัวอย่างขนาดใหญ่เพื่อให้การแจกแจงของสถิติเป็นแบบปกติโดยทฤษฎีนี้กล่าวว่า ถ้าการสุ่มตัวอย่างจากประชากรที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ มีขนาดกลุ่มตัวอย่าง (n) เพิ่มขึ้นแล้ว การแจกแจงของค่าเฉลี่ย \bar{x} ก็จะยิ่งเข้าสู่การแจกแจงของประชากรแบบปกติที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ/\sqrt{n} แต่ในสภาพปัญหาของการวิจัยเชิงทดลองที่ผู้วิจัยไม่อาจที่จะหลีกเลี่ยงต่อการทดลองที่ใช้กลุ่มตัวอย่างขนาดกลาง เช่น 10 , 15 หรือกลุ่มตัวอย่างขนาดเล็ก เช่น 5 ได้ ประการที่หนึ่ง เมื่อปัญหาของการฝ่าฝืนข้อตกลงเป็นสิ่งที่ไม่อาจหลีกเลี่ยง ดังนั้นการสรุปผลของการวิจัยขึ้นอยู่กับ การตัดสินใจของผู้ทดลองในการที่จะเลือกเทคนิคหรือวิธีการสถิติมาทดสอบเพื่อให้ได้ผลสรุปที่เป็นจริง และให้โอกาสของการเกิดความคลาดเคลื่อนต่อการสรุปผลที่เรียกว่า ความคลาดเคลื่อนประเภทที่ 1 เท่ากับที่นักวิจัยระบุไว้ประการที่สองก็คือปัญหาของการใช้เทคนิคการเปรียบเทียบพหุคูณ โดยเฉพาะการเปรียบเทียบภายหลังที่ต้องขึ้นอยู่กับ การทดสอบเอฟ ซึ่งการวิจัยต่างๆได้สนับสนุนและยืนยันว่าเป็นวิธีทดสอบที่มีความแกร่ง แต่ Federer¹ ได้กล่าวว่า เดิมนั้น แม้ว่า การทดสอบเอฟ เพื่อตรวจสอบความมีนัยสำคัญจำเป็นที่จะต้องทำก่อนที่จะวิเคราะห์การเปรียบเทียบพหุคูณ แต่การทดสอบเอฟก็อาจจะเปลี่ยนแปลงคุณสมบัติและผลบางประการของการเปรียบเทียบพหุคูณที่ทำให้ไม่ได้ถูกสำรวจ ประการที่สามด้วยเหตุผลของการที่ผู้วิจัยใช้เทคนิคการเปรียบเทียบพหุคูณวิธีหนึ่งแล้วให้ผลสรุปที่แตกต่างจากเทคนิคการเปรียบเทียบพหุคูณอีกวิธีหนึ่ง

¹ C.W. Dunnett, "Query of Multiple Comparison Tests," Biometrics, (March 1970), pp. 139-142.

บ่อมแสดงให้เห็นว่าได้เกิดผลของความคลาดเคลื่อนประเภทที่ 1 และจากการขาดหลักฐานการศึกษาว่า เทคนิคใดสามารถที่จะควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองได้เท่ากับระดับนัยสำคัญหรืออีกนัยหนึ่งก็คือ อัตราความคลาดเคลื่อนที่ระบุภายใต้สถานการณ์ของการฝ่าฝืนข้อตกลงเบื้องต้นซึ่งเป็นเหตุผลสำคัญของการวิจัยในครั้งนี้ ดังในวัตถุประสงค์ของการวิจัย

วัตถุประสงค์ของการวิจัย

เนื่องจากขาดหลักฐานของการศึกษาในเรื่องผลการฝ่าฝืนข้อตกลงเบื้องต้นของวิธีเปรียบเทียบพหุคูณ การวิจัยครั้งนี้จึงมีจุดมุ่งหมายเพื่อที่จะตรวจสอบผลของการฝ่าฝืนข้อตกลงเบื้องต้นด้านการแจกแจงของประชากร (population distribution) และความแปรปรวนของประชากร (Monte Carlo Simulation) ในการหาผลสรุปลักษณะของอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง (Actual Type I Error Rate) ของวิธีการเปรียบเทียบพหุคูณ 5 วิธี คือ

1. วิธีทดสอบของทูกีย์ (Tukey's HSD Test)
2. วิธีการเปรียบเทียบพหุคูณของดันน์ (Dunn's Multiple Comparisons)
3. วิธีเอสของเชฟเฟย์ (Scheff'e 's S* Method)
4. วิธีทดสอบของดันเนตต์ (Dunnett's Test)
5. วิธีทดสอบของนิวแมนคูลส์ (Newman Keuls Test)

ภายใต้เงื่อนไขดังต่อไปนี้

1. ลักษณะการแจกแจงของประชากรทั้งที่เป็นปกติและไม่เป็นปกติ
2. จำนวนกลุ่มตัวอย่าง 3 กลุ่ม และ 4 กลุ่ม
3. ขนาดกลุ่มตัวอย่างเท่ากัน
4. ความแปรปรวนของประชากรทั้งที่กำหนดให้เท่ากัน และแตกต่างกัน

สมมติฐานของการวิจัย

ผลของการผ่านข้อตกลงเบื้องต้นของการแจกแจงของประชากรและความแปรปรวนของประชากรจะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองของวิธีการเปรียบเทียบพหุคูณทั้ง 5 วิธี แตกต่างกัน

ข้อตกลงเบื้องต้น

1. การวิจัยครั้งนี้ถือว่า อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองเป็นดัชนีสำคัญที่นักวิจัยจะใช้เป็นเกณฑ์ตัดสินเลือกวิธีการเปรียบเทียบพหุคูณ
2. การวิจัยครั้งนี้ไม่คำนึงถึงการคำนวณความมีนัยสำคัญของการทดสอบเอฟ
3. การวิจัยครั้งนี้เลือกใช้โปรแกรมสปรูทที่มีหลักฐานและการศึกษาในการแปลงข้อมูลไปยังประชากรที่ต้องการวิจัย

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ศึกษาตัวแปรดังต่อไปนี้
 - 1.1 ตัวแปรตาม (Dependent variable) คือ อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง (Actual Type I Error Rate) ของเทคนิคการเปรียบเทียบพหุคูณ 5 วิธีที่กล่าวมาแล้วข้างต้น โดยคำนวณเปรียบเทียบกับอัตราความคลาดเคลื่อนที่ระบุ (Norminated α)
 - 1.2 ตัวแปรอิสระ (Independent variable) คือ การแจกแจงของประชากร (population distribution) และความแปรปรวนของประชากร (population variances)
2. ศึกษาผลความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองของเทคนิคต่างๆ เพื่อเปรียบเทียบกับอัตราความคลาดเคลื่อนที่ระบุ เฉพาะเฉพาะการแจกแจงของประชากรที่กำหนดให้ ซึ่งมีการแจกแจงในรูปของการแจกแจงแบบปกติ (normal distribution) การแจกแจงแบบยูนิฟอร์ม (uniform distribution) และการแจกแจงแบบเลปโตเคอร์ติกส์ (Leptokurtic distribution)
3. ศึกษาเฉพาะการเปรียบเทียบค่าเฉลี่ยเป็นรายคู่ (pairwise) เมื่อขนาดกลุ่มตัวอย่างในแต่ละครั้งของการทดลองเท่ากัน

4. กำหนดระดับการทดลองหรือจำนวนกลุ่มตัวอย่าง 3 กลุ่ม และ 4 กลุ่ม ขนาดกลุ่มตัวอย่างในแต่ละระดับการทดลอง คือ 5 , 10 และ 15

5. กำหนดพารามิเตอร์ $\mu = 500$ และ $\sigma^2 = 100$ สำหรับการทดลองเมื่อการแจกแจงและความแปรปรวนเป็นปกติ และจำกัดความแตกต่างของความแปรปรวนเป็น 3 รูปแบบ ดังนี้

5.1 เมื่อความแปรปรวนของประชากรเท่ากัน คือ $\sigma_1^2 = 100, \sigma_2^2 = 100, \sigma_3^2 = 100$ เมื่อจำนวนกลุ่มตัวอย่าง 3 กลุ่ม และ $\sigma_1^2 = 100, \sigma_2^2 = 100, \sigma_3^2 = 100, \sigma_4^2 = 100$ เมื่อจำนวนกลุ่มตัวอย่าง 4 กลุ่ม

5.2 เมื่อความแปรปรวนของประชากรไม่เท่ากันคือ

$$5.2.1 \quad \sigma_1^2 = 90 \quad \sigma_2^2 = 100 \quad \sigma_3^2 = 110$$

$$5.2.2 \quad \sigma_1^2 = 80 \quad \sigma_2^2 = 100 \quad \sigma_3^2 = 120$$

เฉพาะกลุ่มตัวอย่าง 3 กลุ่ม

6. การวิจัยนี้ได้จำลองการทดลองขึ้นในคอมพิวเตอร์ โดยอาศัยเทคนิคที่เรียกว่ามอนติคาร์โล ซิมูเลชัน (Monte Carlo Simulation Technique) และโปรแกรมสำเร็จรูป โดยจำกัดการศึกษาเฉพาะการแจกแจงของประชากรใน 3 รูปแบบ คือ การแจกแจงแบบปกติ การแจกแจงแบบยูนิฟอร์ม และการแจกแจงแบบเลปโตเคอร์ติคัล

ตาราง แสดงค่าสถิติของการแจกแจงประชากร เมื่อกลุ่มตัวอย่างขนาด 10,000 ตัว

Population Distribution	μ	σ^2	\bar{X}	S^2	Kurtosis	Skewness
Normal	500	100	500	100	1.840	0.162
Uniform	500	100	500	100	1.914	0.465
Leptokurtics	500	100	499	100	2.026	0.452

7. ศึกษาเฉพาะกลุ่มตัวอย่างขนาดกลางเมื่อกำหนดอัตราความคลาดเคลื่อนใน 2 ระดับ คืออัตราความคลาดเคลื่อนที่ระบุในระดับ .05 และระดับ .01 เปรียบเทียบกับอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง

8. ศึกษาเฉพาะวิธีการเปรียบเทียบพหุคูณ 5 วิธี คือ วิธีทดสอบของทูกี วิธีเปรียบเทียบพหุคูณของดันทน์ วิธีเอสของเชฟเฟย์ วิธีทดสอบของดันทน์เนตต์และวิธีทดสอบของนิวแมนคูลล์

9. การชิวมูล การทดลองนี้จะกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง เพื่อประหยัดเวลาในการทำงานของคอมพิวเตอร์

คำจำกัดความ

ความแกร่ง หมายถึง คุณสมบัติของการทดสอบหาสถิติที่ไม่แสดงถึงความไว (sensitive) ต่อการเบี่ยงเบนหรือการฝ่าฝืนข้อตกลง เบื้องต้นของการทดสอบที่มีผลต่ออัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 และความคลาดเคลื่อนประเภทที่ 2

ความคลาดเคลื่อนประเภทที่ 1 หมายถึง โอกาสที่จะปฏิเสธสมมติฐานที่ตั้งไว้เมื่อสมมติฐานนั้นเป็นจริง โอกาสที่จะเกิดความผิดพลาดประเภทนี้ กำหนดด้วยอัตราความคลาดเคลื่อนที่ระบุในระดับ α

อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองหมายถึงอัตราความคลาดเคลื่อนประเภทที่ 1 ที่เกิดขึ้นจริงในการทดสอบสมมติฐานของการทดลองตามแผนการทดลอง

อัตราความคลาดเคลื่อนที่ระบุ หมายถึง อัตราความคลาดเคลื่อนประเภทที่ 1 ที่ควบคุมด้วยระดับนัยสำคัญ α และใช้เป็นเกณฑ์ในการตัดสินอัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลองของวิธีเปรียบเทียบพหุคูณทั้ง 5 วิธีที่ใช้ในการวิจัยครั้งนี้

ประโยชน์ของการวิจัย

เพื่อช่วยให้นักวิจัย หรือผู้ใช้สถิติได้มั่นใจในการเลือกเทคนิควิธีการเปรียบเทียบพหุคูณแบบต่างๆ ให้เหมาะสมกับข้อมูลที่น่ามาพิจารณาตัดสิน อันจะทำให้ได้ผลลัพธ์ที่เกิดความคลาดเคลื่อนในทางสถิติที่เรียกว่าความคลาดเคลื่อนประเภทที่ 1 เท่ากับระดับนัยสำคัญที่กำหนด