

บทที่ 1

บทนำ

ความเป็นมาของปัญหา

การใช้วิธีการทางสถิติในการเปรียบเทียบความเท่ากันของค่าเฉลี่ยประชากร 2 กลุ่ม เพื่อตอบปัญหาในชีวิตประจำวัน ตัวอย่างเช่น เจ้าของโรงงานต้องการทราบว่าเครื่องจักรที่ใช้ในการผลิต 2 เครื่องมีประสิทธิภาพในการผลิตเท่ากันหรือไม่ สามารถกระทำได้โดยการเปรียบเทียบผลผลิตเฉลี่ยที่ได้จากเครื่องจักรแต่ละเครื่องว่าให้ผลผลิตแตกต่างกันหรือไม่ภายใต้เงื่อนไขเดียวกัน ครูต้องการทราบว่าผลการเรียนวิชาคณิตศาสตร์ของนักเรียน 2 ห้องแตกต่างกันหรือไม่ โดยวิธีการเปรียบเทียบคะแนนสอบเฉลี่ยของนักเรียนทั้ง 2 ห้องว่าแตกต่างกันหรือไม่ ในการวางแผนการจ้างงานผู้บริหารต้องทำการตัดสินใจว่าจะเลือกจ้างลูกจ้างประจำหรือลูกจ้างชั่วคราวดีกว่ากัน โดยใช้วิธีการเปรียบเทียบผลผลิตเฉลี่ยที่ได้จากลูกจ้างทั้ง 2 ประเภทว่าแตกต่างกันหรือไม่ เป็นต้น ซึ่งการตอบปัญหาดังกล่าวข้างต้นสามารถกระทำได้โดยใช้วิธีการทดสอบสมมติฐานทางสถิติ

ในการทดสอบสมมติฐานสำหรับงานวิจัยต่าง ๆ ผู้วิจัยต้องทำการออกแบบงานวิจัยและเลือกสถิติทดสอบสำหรับการทดสอบให้เหมาะสมกับข้อมูลที่ต้องการวิเคราะห์ ซึ่งตัวสถิติที่ใช้สำหรับการทดสอบแต่ละวิธีต่างก็มีข้อตกลงเบื้องต้นไม่เหมือนกัน ดังนั้น ผู้วิจัยจะต้องมีความสามารถในการวางแผนการตอบปัญหาที่ดี มีความเข้าใจในเรื่องที่กำลังศึกษาและสามารถที่จะเลือกสถิติทดสอบได้เหมาะสมด้วย จึงจะทำให้ผลสรุปของการวิจัยมีความถูกต้องและน่าเชื่อถือมากขึ้น

การสรุปผลเชิงสถิติเกี่ยวกับประชากรโดยทั่วไปอาจกระทำได้โดยอาศัยตัวอย่างที่เลือกโดยวิธีการสุ่มมาจากประชากรที่ทราบว่ามิลักษณะการแจกแจงอย่างใดอย่างหนึ่ง ซึ่งประสิทธิภาพของผลสรุปจะขึ้นอยู่กับเงื่อนไขข้อตกลงเบื้องต้นเกี่ยวกับลักษณะการแจกแจงของประชากรและค่าความแปรปรวน ดังนั้นในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร 2 กลุ่ม ซึ่งเป็นการทดสอบที่ใช้พารามิเตอร์ (parametric test) จะอยู่ภายใต้ข้อตกลงเบื้องต้นที่ว่า ตัวอย่างทั้งสองกลุ่มที่นำมาทดสอบต้องสุ่มมาจากการแจกแจงปกติและมีค่าความแปรปรวนคงที่และเท่ากัน แต่ในสถานการณ์โดยทั่วไปนั้นลักษณะของข้อมูลที่ได้อาจไม่สอดคล้องกับข้อตกลงเบื้องต้นของตัวสถิติทดสอบที่ใช้พารามิเตอร์ นับเป็นปัญหาสำคัญที่ผู้วิจัยโดยส่วนใหญ่ประสบในการเลือกสถิติทดสอบให้เหมาะสมกับข้อมูลที่ได้อัน ซึ่งในกรณีเช่นนี้ผู้วิจัยอาจใช้วิธีการแปลงข้อมูล (Transformation of data) เพื่อให้ข้อมูลที่ได้อันเป็นไปตามข้อตกลงเบื้องต้น แล้วจึงนำข้อมูลที่แก้ไขปรับปรุงแล้วมาทำการวิเคราะห์ (ซึ่งไม่ค่อยเป็นที่นิยมเนื่องจากมีปัญหาในการเลือกวิธีการแปลงข้อมูลที่เหมาะสมและมีขั้นตอนที่ยุ่งยากในบางกรณีไม่สามารถแปลงข้อมูลให้เป็นไปตามข้อตกลงเกี่ยวกับค่าความแปร

ปรวนที่ต้องเท่ากันได้) หรืออาจใช้วิธีการทดสอบที่ไม่ใช้พารามิเตอร์ (Nonparametric test) ซึ่งเป็นการทดสอบที่ไม่ขึ้นอยู่กับลักษณะการแจกแจงของประชากร (Distribution free) และมีวิธีการที่ง่ายต่อการคำนวณและการนำไปใช้

สถิติทดสอบที่ใช้พารามิเตอร์ในการทดสอบสมมติฐานเกี่ยวกับความเท่ากันของค่าเฉลี่ยประชากร 2 กลุ่มที่เป็นอิสระกัน โดยกำหนดให้ μ_1 และ μ_2 แทนค่าเฉลี่ยของประชากรกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ มักนิยมใช้สถิติทดสอบ Z-test หรือ t-test ซึ่งจะขึ้นอยู่กับลักษณะข้อมูลที่นำมาใช้ทดสอบดังนี้

1. กรณีที่ทราบหรือไม่ทราบค่าความแปรปรวนของประชากร แต่ขนาดตัวอย่างในแต่ละกลุ่มประชากรมากกว่า 30 จะใช้สถิติ Z-test ซึ่งเป็นการแจกแจงปกติมาตรฐานที่มีค่าเฉลี่ย เป็น 0 และค่าความแปรปรวนเป็น 1

2. กรณีที่ไม่ทราบค่าความแปรปรวนของประชากร และขนาดตัวอย่างในแต่ละกลุ่มน้อยกว่าหรือเท่ากับ 30 จะใช้สถิติ t-test

- กรณีที่ทราบว่าประชากร 2 กลุ่มมีความแปรปรวนเท่ากัน สถิติ pooled t-test มีการแจกแจงแบบที ที่องศาความเป็นอิสระเท่ากับ n_1+n_2-2 นั่นคือขนาดตัวอย่างกลุ่มที่ 1 (แทนด้วย n_1) บวกกับขนาดตัวอย่างกลุ่มที่ 2 (แทนด้วย n_2) ลบด้วย 2

- กรณีที่ทราบว่าประชากร 2 กลุ่มมีความแปรปรวนที่แตกต่างกัน (Behrens-Fisher Problem) นั้น ได้มีผู้นำเสนอสถิติเพื่อใช้ในการนี้ดังกล่าว อาทิเช่น เวลช์ (Welch, 1937, 1947) ได้เสนอสถิติ t-test (Welch Approximate t-test) โดยการประมาณองศาความเป็นอิสระ (df) จากค่าความแปรปรวนของกลุ่มตัวอย่างทั้งสองกลุ่ม คือ

$$df = \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\sqrt{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}}}$$

โดยที่ s_1^2 และ s_2^2 เป็นค่าความแปรปรวนของตัวอย่างสุ่มจากประชากรกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ ซึ่งสามารถหาค่าวิกฤติได้จากตารางการแจกแจงที, เซทเทอร์ทไวท์ (Satterthwaite, 1946) เสนอสถิติทดสอบค่าเฉลี่ยประชากร 2 กลุ่มที่ค่าความแปรปรวนแตกต่างกัน ในรูปของสถิติ t^2 -test โดยการประมาณค่าองศาความเป็นอิสระจากค่าความแปรปรวนของตัวอย่างเช่นเดียวกับของเวลช์ แต่จะหาค่าวิกฤติได้จากตารางการแจกแจงเอฟ ซึ่งให้ผลในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และกำลังการทดสอบที่ยังไม่คืนักในกรณีที่การแจกแจงไม่ใช่การแจกแจงแบบปกติ และในปี ค.ศ.1948 เวลช์และเอสปิน (Welch and Aspin, 1948) ได้นำเสนอสถิติทดสอบ t-test ที่คำนวณค่าวิกฤติได้จากฟังก์ชันที่ขึ้นกับขนาดตัวอย่างและตารางการแจกแจงที แต่มีขั้นตอนการคำนวณที่ค่อนข้างยุ่งยากจึงไม่มีที่นิยม จึงได้มีผู้นำเสนอสถิติ

ทดสอบที่มีขั้นตอนการคำนวณค่าวิกฤติที่ง่ายขึ้นและมีความสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี ทั้งยังให้กำลังการทดสอบที่สูงขึ้นด้วย ได้แก่ ลีและเกอร์แลนด์ (Lee F.S. A. and Gurland John; 1975) โดยการเสนอรูปแบบฟังก์ชันในการคำนวณค่าวิกฤติ ซึ่งขึ้นอยู่กับขนาดตัวอย่างและความแปรปรวนของกลุ่มตัวอย่างทั้งสองกลุ่ม เรียกว่า สถิติทดสอบลี-เกอร์แลนด์ (Lee-Gurland test) จากนั้น ลี (Lee F.S.A., 1995) ได้เสนอสถิติ Fitted Test (ซึ่งพัฒนามาจากสถิติทดสอบลี-เกอร์แลนด์ Lee-Gurland test, 1975) โดยการประมาณค่าสัมประสิทธิ์ในการคำนวณค่าวิกฤติของรูปแบบฟังก์ชันของลี-เกอร์แลนด์ เพื่อการคำนวณที่สะดวกขึ้น และลี (Lee) ยังพบว่าสถิติทดสอบที่ได้มีอำนาจการทดสอบสูงกว่าสถิติทดสอบเวลช์ที่และสถิติทดสอบลีเกอร์แลนด์เดิม และมีความแกร่ง (Robustness) เมื่อประชากรมีลักษณะการแจกแจงที่ไม่ใช่การแจกแจงแบบปกติ

ส่วนการทดสอบที่ไม่ใช่พารามิเตอร์ได้มีการพัฒนาสถิติทดสอบสมมติฐานเกี่ยวกับค่ากลางของประชากรสองกลุ่มที่เป็นอิสระกัน ในกรณีที่มีการแจกแจงของประชากรไม่ใช่การแจกแจงแบบปกติ และมีวิธีการคำนวณที่ง่าย โดยได้มีผู้เสนอแบบทดสอบไว้และก็ได้มีการพัฒนาขึ้นมาอีกมาก อาทิเช่น ไมเคิล เอ ฟลิเกอร์ และ จอร์จ อี พอลิเซลโล (Michael A. Fligner and George E. Policello, 1981) โดยเสนอการทดสอบแมนวิทนียี่วิลคอกชันส่วนขยาย ที่พัฒนามาจากการทดสอบแมนวิทนียี่วิลคอกชัน ซึ่งมีกำลังการทดสอบเท่ากับการทดสอบแมนวิทนียี่วิลคอกชันเดิม แต่จะสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า เมื่อประชากรมีค่าความแปรปรวนไม่เท่ากัน และ มีการแจกแจงแตกต่างกันด้วย และ เอช เอช เลมเมอร์ (H.H. Lemmer, 1987) ได้เสนอการทดสอบแมนวิทนียี่วิลคอกชันส่วนขยาย ที่พัฒนามาจากการทดสอบแมนวิทนียี่วิลคอกชัน ซึ่งมีความแกร่งมากกว่าการทดสอบแมนวิทนียี่วิลคอกชันเดิม ในกรณีที่มีการแจกแจงของประชากรไม่ใช่การแจกแจงแบบปกติและมีความแปรปรวนของประชากรแตกต่างกัน และไม่ทราบค่า ซึ่งดวงแก้ว ปรีดีสนิท (พ.ศ. 2534) ได้อ้างถึงสถิติทดสอบแมนวิทนียี่วิลคอกชันส่วนขยายทั้ง 2 วิธี โดยได้ทำการเปรียบเทียบกับสถิติทดสอบที่และสถิติทดสอบสตีเวนที่วิลคอกชัน ภายใต้ลักษณะการแจกแจงของประชากรเป็นแบบสมมาตรหางยาว ที่ขนาดตัวอย่าง 4 และ 8 พบว่าสถิติแมนวิทนียี่วิลคอกชันส่วนขยายทั้ง 2 วิธี มีประสิทธิภาพในการทดสอบสูงกว่าสถิติทดสอบที่และสตีเวนที่วิลคอกชัน

ดังนั้นจึงเป็นที่น่าสนใจศึกษาต่อไปว่าสถิติทดสอบ 3 วิธี ได้แก่ สถิติฟิตเต็ดของลี (1995), สถิติแมนวิทนียี่วิลคอกชันส่วนขยายทั้ง 2 วิธี ภายใต้ลักษณะการแจกแจงประชากรที่ไม่ใช่การแจกแจงแบบปกติ เช่น การแจกแจงแบบเบ้ ว่าสถิติทดสอบใดมีอำนาจการทดสอบสูงกว่ากัน หรือมีความเหมาะสมกับการนำไปใช้ในสถานการณ์ต่าง ๆ ได้ดีกว่ากัน

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบอำนาจการทดสอบของสถิติทดสอบค่าเฉลี่ยประชากร 2 กลุ่มที่เป็นอิสระกัน ได้แก่ การทดสอบฟิตเต็ด (Fitted Test : FT) การทดสอบแมนวิทนียี่วิลคอกชันส่วนขยาย (Modified

Mann-Whitney Wilcoxon : \hat{U} ที่เสนอโดยไมเคิล เอ ฟลิทเนอร์ และ จอร์จ อี พอลิเชลโล, 1981) และการทดสอบแมนวิทนียวีลคอกชันส่วนขยาย (Modified Mann-Whitney Wilcoxon : SU ที่เสนอโดย เอช เอช เลมเมอร์, 1987) เมื่อประชากรทั้งสองกลุ่มมีค่าความแปรปรวนแตกต่างกันและไม่ทราบค่า และมีลักษณะการแจกแจงของประชากรไม่ใช่การแจกแจงแบบปกติ (แบบเบ้) โดยศึกษาเปรียบเทียบภายใต้สถานการณ์ต่าง ๆ ตามที่กำหนดไว้ในขอบเขตของการวิจัย เพื่อนำเสนอแนวทางการเลือกใช้สถิติทดสอบค่าเฉลี่ยประชากร 2 กลุ่ม ให้เหมาะสมกับลักษณะของข้อมูลที่จะนำมาใช้ในการวิเคราะห์ทางสถิติเพื่อตอบปัญหาการวิจัย

สมมติฐานการวิจัย

ภายใต้ลักษณะการแจกแจงของประชากรที่ไม่ใช่การแจกแจงแบบปกติ (แบบเบ้) ค่าความแปรปรวนของประชากร และขนาดตัวอย่างที่แตกต่างกัน มีผลทำให้กำลังการทดสอบของแบบทดสอบค่าเฉลี่ยประชากร 2 กลุ่มที่ทำการศึกษามีความแตกต่างกัน

ข้อตกลงเบื้องต้น

การเปรียบเทียบกำลังการทดสอบของแบบทดสอบในการวิจัยครั้งนี้ จะทำการศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ก่อน แล้วจึงจะทำการศึกษาเปรียบเทียบกำลังการทดสอบของสถิติทดสอบ เพื่อเลือกสถิติทดสอบที่มีกำลังการทดสอบสูงสุด

ขอบเขตของการวิจัย

1. ศึกษาเปรียบเทียบสถิติทดสอบค่าเฉลี่ยประชากร 2 กลุ่ม เมื่อค่าความแปรปรวนของประชากรทั้งสองกลุ่มแตกต่างกันและไม่ทราบค่า ดังนี้

1.1 การทดสอบฟิตเต็ด (Fitted Test : FT)

1.2 การทดสอบแมนวิทนียวีลคอกชันส่วนขยาย (Modified Mann-Whitney Wilcoxon : \hat{U} ที่เสนอโดยไมเคิล เอ ฟลิทเนอร์ และ จอร์จ อี พอลิเชลโล, 1981)

1.3 การทดสอบแมนวิทนียวีลคอกชันส่วนขยาย (Modified Mann-Whitney Wilcoxon : SU ที่เสนอโดย เอช เอช เลมเมอร์, 1987)

2. กำหนดลักษณะการแจกแจงของประชากรเป็นแบบเบ้ ได้แก่ การแจกแจงแกมมา และการแจกแจงไวบูลล์

3. กำหนดขนาดตัวอย่างที่ศึกษา คือ 6 และ 10 โดยที่ $n_1 \geq n_2$

4. กำหนดความแตกต่างของค่าเฉลี่ยประชากรของทั้งสองกลุ่ม ($\mu_1 - \mu_2 = d$) เป็น 0, 2, 5 และ 20
5. กำหนดสัดส่วนความแปรปรวนของประชากรกลุ่มที่ 1 ต่อ ความแปรปรวนของประชากรกลุ่มกลุ่มที่ 2 ($R = \frac{\sigma_1^2}{\sigma_2^2}$) เป็น 1.5, 2, 3 และ 5
6. กำหนดระดับนัยสำคัญของการทดสอบเป็น 2 ระดับ คือ 0.01 และ 0.05
7. ข้อมูลที่ใช้ในการศึกษาจะใช้วิธีการจำลองข้อมูลให้มีสถานการณ์ตามที่กำหนด และทดสอบสมมติฐานโดยการกระทำซ้ำ 100 ครั้งในแต่ละสถานการณ์

คำจำกัดความ

ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) หมายถึง ความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างนั้นจริง และความน่าจะเป็นของความผิดพลาดชนิดนี้คือ ระดับนัยสำคัญของการทดสอบ (α)

ความคลาดเคลื่อนประเภทที่ 2 (Type II Error) หมายถึง ความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่างเมื่อสมมติฐานว่างนั้นไม่จริง ความน่าจะเป็นของความผิดพลาดชนิดนี้แทนด้วย β

กำลังการทดสอบ (Power of The Test) หมายถึง ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างนั้นไม่จริง ซึ่งมีค่าเท่ากับ $1 - \beta$ เมื่อ β คือ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2

ความแกร่ง (Robustness) หมายถึง คุณสมบัติของการทดสอบที่ไม่แสดงความไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น กรณีที่ลักษณะของข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นของการทดสอบ เกี่ยวกับการแจกแจงหรือค่าความแปรปรวนที่เปลี่ยนแปลงไป อันจะมีผลกระทบต่อความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (α) และ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 (β) แต่การทดสอบที่มีความแกร่งนั้น ยังคงให้ผลสรุปที่ไม่แตกต่างไปจากผลการสรุปเดิม

ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางให้ผู้วิจัยสามารถเลือกใช้สถิติทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร 2 กลุ่ม กรณีที่ค่าความแปรปรวนของประชากรทั้งสองกลุ่มแตกต่างกันและไม่ทราบค่า และเมื่อข้อมูลมีลักษณะการแจกแจงไม่เป็นไปตามข้อตกลงเบื้องต้นที่ว่าข้อมูลตัวอย่างต้องมาจากประชากรที่มีการแจกแจงปกติ