ว.วิทย. มข. 41(3) 638-647 (2556) KKU Sci. J. 41(3) 638-647 (2013)

การเปรียบเทียบสถิติทดสอบโดยใช้การจำลองข้อมูล

Comparison of Test Statistics Using Simulated Data

มานะชัย รอดชื่น[1]

บทคัดย่อ

บทความนี้ได้พิจารณาการเปรียบเทียบสถิติทดสอบโดยใช้การจำลองข้อมูล ภายใต้สถิติทดสอบที่มีขนาด α หรือ มีระดับ α เกณฑ์ที่ใช้ในการจัดกลุ่มการทดสอบได้ใช้เกณฑ์ของ Cochran (1947) เกณฑ์ของ Bradley (1978, cited in Tomarken and Serlin, 1986) และเกณฑ์การพิจารณาภายใต้การทดสอบสมมติฐานของค่า α ซึ่งพิจารณาการแจกแจงของสถิติทดสอบสองแบบคือประมาณการแจกแจงของสถิติทดสอบด้วยการแจกแจงปรกติและใช้การแจกแจงทวินาม

ABSTRACT

This article considers the comparison of the test statistics under the test of size α or level α. The criteria of Cochran (1947), the criteria of Bradley (1978, cited in Tomarken and Serlin, 1986) and the criteria under hypothesis testing for α are applied with distributions that are well approximated by a normal distribution and binomial distribution.

คำสำคัญ: สถิติทดสอบ การจำลอง ขนาดของการทดสอบ การทดสอบสมมติฐาน

Keywords: Test statistics, Simulation, Size of the test, Hypothesis testing

[1]ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ อำเภอเมือง จังหวัดเชียงใหม่ 50200

E-mail: r.manachai@gmail.com

1. บทนำ

ในการประมาณค่าพารามิเตอร์ การพิจารณาตัวประมาณที่ดีจะดูจากคุณสมบัติต่าง ๆ ของตัวประมาณ เช่น ความไม่เอนเอียง (unbiased) ความต้องกัน (consistency) ความแปรปรวน (variance) หรือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (mean squared error) ส่วนการทดสอบสมมติฐาน การเลือกสถิติทดสอบที่เหมาะสมส่วนใหญ่จะพิจารณาจากขนาดของการทดสอบ (size of the test) หรือ กำลังของการทดสอบ (power of the test)

สำหรับการทดสอบสมมติฐานจะพบว่าการหาขนาดของการทดสอบ หรือกำลังของการทดสอบนั้นทำได้ไม่ง่ายในบางกรณี เนื่องจากอาจจะเกิดจากปัญหา เช่น ตัวแปรที่สนใจศึกษาไม่ได้มีการแจกแจงปรกติ การเกิดอัตตสหสัมพันธ์ ข้อมูลที่สูญหายไป หรือกรณีข้อมูลมีค่าผิดปกติ ซึ่งได้มีนักวิชาการเสนอแนวคิดในการแก้ปัญหาดังกล่าวเป็นจำนวนมาก แต่การแก้ปัญหามักใช้ได้ภายใต้ข้อจำกัดบางประการเท่านั้น สำหรับการเลือกวิธีการแก้ปัญหาที่เหมาะสมกับสถานการณ์ต่าง ๆ โดยทั่วไปจะใช้ผลการจำลองข้อมูลมาพิจารณาเป็นเบื้องต้น

ในปัจจุบันมีวารสารทางวิชาการหลาย ๆ เรื่องที่ใช้ผลการจำลองข้อมูลพิจารณาร่วมกับตัวสถิติทดสอบที่สร้างขึ้นในการทดสอบสมมติฐาน สุพรรณี (2541) ได้แสดงขั้นตอนในการจำลองข้อมูลเพื่อหาตัวประมาณ และทดสอบสมมติฐาน Rizzo (2008) ได้แสดงแนวคิดในการจำลองโดยใช้โปรแกรม R เป็นต้น

สำหรับการจำลองข้อมูลเพื่อเปรียบเทียบสถิติทดสอบจะพิจารณาจากค่าประมาณขนาดของการทดสอบ $(\hat{\alpha})$ หรือค่าประมาณกำลังของการทดสอบ $(1-\hat{\beta})$ หรือพิจารณาทั้งสองอย่างควบคู่กัน ซึ่งสำหรับการพิจารณาทั้งสองอย่างควบคู่กันนั้น อันดับแรกจะดูที่ค่าประมาณขนาดของการทดสอบก่อนว่าอยู่ในเกณฑ์ที่กำหนดหรือไม่ (เรียกว่าการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1) หากอยู่ในเกณฑ์ที่กำหนด จากนั้นจึงเลือกสถิติทดสอบที่ให้ค่าประมาณกำลังของการทดสอบมากที่สุดเป็นสถิติทดสอบที่เหมาะสม

สำหรับเกณฑ์ในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ที่นิยมใช้ ได้แก่ เกณฑ์ของ Cochran (1947) เกณฑ์ของ Bradley (1978, cited in Tomarken and Serlin, 1986) และเกณฑ์การพิจารณาภายใต้การทดสอบสมมติฐานของค่า α ที่ประมาณการแจกแจงของสถิติทดสอบด้วยการแจกแจงปรกติ

สำหรับบทความนี้ได้ชี้ให้เห็นการใช้เกณฑ์การพิจารณาภายใต้การทดสอบสมมติฐานของค่า α ซึ่งสถิติทดสอบมีการแจกแจงทวินาม และนำเสนอพื้นฐานของหลักการทดสอบสมมติฐาน ความคลาดเคลื่อนในการทดสอบสมมติฐาน ความสัมพันธ์ระหว่างความคลาดเคลื่อนชนิดที่ 1 และชนิดที่ 2 ซึ่งเป็นการแสดงความสัมพันธ์ของขนาด และกำลังของการทดสอบนั่นเอง

2. การทดสอบสมมติฐาน

การทดสอบสมมติฐาน (hypothesis testing) เป็นการทดสอบค่าพารามิเตอร์ของประชากรว่าเป็นไปตามที่คาดหรือไม่ (การตัดสินใจเกี่ยวกับค่าของพารามิเตอร์) การกำหนดสมมติฐานนั้นมีการกำหนด 2 แบบ คือ สมมติฐานเชิงบรรยาย (descriptive hypothesis) ซึ่งเป็นสมมติฐานที่อยู่ในลักษณะของข้อความการบรรยาย และสมมติฐานเชิงสถิติ (statistical hypothesis) คือ “ข้อความ (statement) หรือข้อเสนอ (assertion) หรือ

ประพจน์ (proposition) เกี่ยวกับการแจกแจงความน่าจะเป็นของตัวแปรสุ่มตัวหนึ่งหรือหลายตัว” (ประชุม, 2545) ซึ่งนิยมเขียนอยู่ในรูปของสัญลักษณ์ทางคณิตศาสตร์ของพารามิเตอร์ สมมติฐานแบ่งเป็น 2 ชนิด คือ สมมติฐานจริง (สมมติฐานเพื่อการทดสอบ) หรือสมมติฐานว่าง (null hypothesis) แทนด้วย H_0 และสมมติฐานแย้ง (alternative hypothesis) แทนด้วย H_1

ในการทดสอบสมมติฐานจะใช้ตัวอย่างสุ่ม $X_1, X_2, ..., X_n$ ที่อยู่ในวงศ์ของการแจกแจง $\mathcal{P} = \{f(x,\theta), \theta \in \Omega\}$ โดยที่ Ω แทนปริภูมิพารามิเตอร์ และกำหนด $\Omega_0 \subset \Omega, \Omega_1 \subset \Omega$ เมื่อ Ω_0, Ω_1 คือปริภูมิพารามิเตอร์ภายใต้ H_0 และ H_1 ตามลำดับ ซึ่ง $\Omega_0 \cap \Omega_1 = \phi$ (empty set) และวงศ์ของการแจกแจงที่สอดคล้องกับ Ω_0, Ω_1 จะแทนด้วย $\mathcal{P}_0$, $\mathcal{P}_1$ ตามลำดับ กำหนดผลแบ่งกั้นของปริภูมิตัวอย่าง S แทนด้วย $\{S_0, S_1\}$ การตัดสินใจจะขึ้นกับ S ถ้าค่าของข้อมูลจากตัวอย่างสุ่ม $X_1, X_2, ..., X_n$ เป็นสมาชิกของ S_0 $(\underset{\sim}{x} \in S_0)$ สมมติฐานภายใต้ $\mathcal{P}_0$ ก็ไม่สามารถที่จะปฏิเสธได้ และถ้า $\underset{\sim}{x} \in S_1$ ก็ไม่สามารถที่จะยอมรับ $\mathcal{P}_0$ ได้ และจะเรียก S_1 ว่าบริเวณปฏิเสธสมมติฐาน หรือบริเวณวิกฤต (region of rejection or critical region) (Lehmann, 1986)

โดยทั่วไปแล้วการตัดสินใจยอมรับหรือปฏิเสธสมมติฐานจะพิจารณาจากการแจกแจงของตัวสถิติที่ได้จากตัวอย่าง (ขึ้นอยู่กับวิธีการหาการทดสอบ) ตัวอย่างเช่น ในการประมาณค่าเฉลี่ยของประชากร μ นั้น โดยปกติแล้วเราจะใช้ค่าเฉลี่ยของตัวอย่าง $\bar{X}$ ทำการประมาณ ทำนองเดียวกันในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร μ ก็จะพิจารณาจากการแจกแจงของค่าเฉลี่ยของตัวอย่าง $\bar{X}$ จะเรียกว่า **สถิติเพื่อการทดสอบ** (test statistic) การที่จะยอมรับหรือปฏิเสธสมมติฐานนั้น จะพิจารณาจากค่าสถิติเพื่อการทดสอบว่าตกอยู่ในอาณาเขตที่ยอมรับ H_0 หรือ ปฏิเสธ H_0 และจะเรียกอาณาเขตปฏิเสธ H_0 ว่า **อาณาเขตวิกฤต** (critical region) ซึ่งสอดคล้องกับ S_1 ส่วนค่าที่กำหนดขอบเขตของอาณาเขตปฏิเสธก็จะเรียกว่า **ค่าวิกฤต** (critical value)

3. ความคลาดเคลื่อนในการทดสอบสมมติฐาน

เนื่องจากการใช้ข้อมูลจากตัวอย่างอ้างอิงถึงประชากรอาจทำให้เกิดความคลาดเคลื่อนในการสรุปผล ซึ่งความคลาดเคลื่อนมี 2 ชนิด

1) ความคลาดเคลื่อนชนิดที่ 1 (type I error) คือ ความคลาดเคลื่อนที่เกิดจากการตัดสินใจปฏิเสธ H_0 โดยที่ H_0 เป็นจริง และขนาดของความคลาดเคลื่อนชนิดที่ 1 (size of type I error) คือ ความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนชนิดนี้ แทนด้วย $P(\text{ ปฏิเสธ } H_0 | H_0 \text{ จริง }) = P(\underset{\sim}{X} \in S_1 | \theta \in \Omega_0) = \alpha$ และระดับนัยสำคัญ (level of significance) คือ $P(\underset{\sim}{X} \in S_1) \leq \alpha, \forall \theta \in \Omega_0$

2) ความคลาดเคลื่อนชนิดที่ 2 (type II error) คือ ความคลาดเคลื่อนที่เกิดจากการตัดสินใจยอมรับ H_0 โดยที่ H_0 ไม่จริง และขนาดของความคลาดเคลื่อนชนิดที่ 2 (size of type II error) คือ ความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนชนิดนี้แทนด้วย β นั่นคือ $P(\text{ ยอมรับ } H_0 | H_0 \text{ ไม่จริง}) = \beta$ และจะเรียก $1-\beta$ ว่า **กำลังของการทดสอบ** (power of the test) ซึ่งเป็นการตัดสินใจถูกต้องที่จะปฏิเสธ H_0 โดยที่ H_0 ไม่จริง ในการสร้างหรือพิจารณาสถิติทดสอบที่เหมาะสมจะควบคุมให้ α มีค่ามากที่สุดที่ยอมให้เกิดขึ้นได้ และทำให้ β มีค่าน้อยที่สุดเพราะจะทำให้กำลังของการทดสอบมีค่ามากที่สุดนั่นเอง (ธีระพร, 2536; Lehmann, 1986) ในการเกิดความ

คลาดเคลื่อนจะเกิดเฉพาะชนิดใดชนิดหนึ่งเท่านั้น แต่เนื่องจากไม่ทราบว่าสมมติฐานที่ทดสอบเป็นจริงหรือไม่ ความคลาดเคลื่อนดังกล่าวจึงเป็นการกำหนดขึ้นภายใต้สถานการณ์ $\theta \in \Omega_0$ หรือสถานการณ์ $\theta \in \Omega_1$ โดยที่ $\Omega_0 \cap \Omega_1 = \phi$

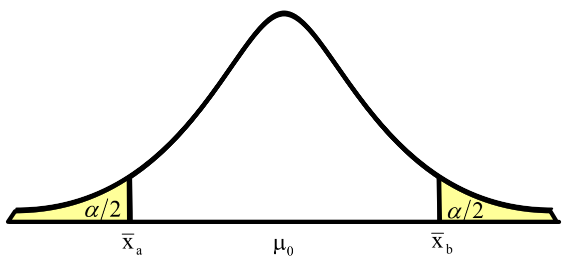
4. ความสัมพันธ์ระหว่างความคลาดเคลื่อนชนิดที่ 1 และชนิดที่ 2

เนื่องจากการเกิดความคลาดเคลื่อนชนิดที่ 1 และ 2 จะเกิดขึ้นอย่างใดอย่างหนึ่งเท่านั้น และผลจากการทดสอบสมมติฐานก็มีสองทางเลือกที่เป็นไปได้ ดังนั้นถ้ายอมรับสมมติฐาน H_0 ก็จะทำให้อาจเกิดความคลาดเคลื่อนชนิดที่ 2 ได้ เมื่อ H_0 ไม่จริง ทำนองเดียวกันถ้าปฏิเสธสมมติฐาน H_0 ก็อาจทำให้เกิดความคลาดเคลื่อนชนิดที่ 1 ได้ เมื่อ H_0 เป็นจริง อีกทั้งค่าความน่าจะเป็นของความคลาดเคลื่อนทั้งสองชนิดจะแปรผกผันกัน ซึ่งจะขึ้นอยู่กับบริเวณวิกฤต โดยทั่วไปในการทดสอบสมมติฐานจะหาค่าวิกฤต จากการกำหนดระดับนัยสำคัญ (α) ไว้ล่วงหน้า และตัวสถิติทดสอบ ความสัมพันธ์ระหว่างค่า α และ β จะแสดงให้เห็นดังตัวอย่างต่อไปนี้

ตัวอย่าง ในการทดสอบสมมติฐาน $H_0 : \mu = \mu_0$ เทียบกับ $H_1 : \mu \neq \mu_0 (\mu = \mu_1)$ กรณีที่ตัวอย่างสุ่มได้มาจากประชากรที่มีการแจกแจงปรกติ และทราบค่าความแปรปรวน σ^2 กำหนดระดับนัยสำคัญที่ทดสอบเท่ากับ α

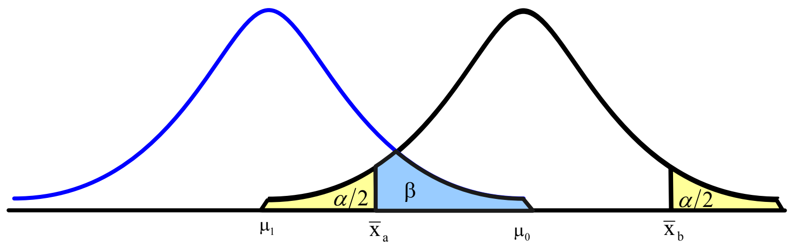
ดังนั้นตัวสถิติทดสอบ คือ $Z = \dfrac{\bar{X} - \mu_o}{\sigma/\sqrt{n}}$ มีบริเวณวิกฤต คือ $z_{cal} \leq -z_{1-\frac{\alpha}{2}}$ หรือ $z_{cal} \geq z_{1-\frac{\alpha}{2}}$,

$\bar{x} \leq \bar{x}_a = \mu_0 - z_{1-\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$ หรือ $\bar{x} \geq \bar{x}_b = \mu_0 + z_{1-\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$

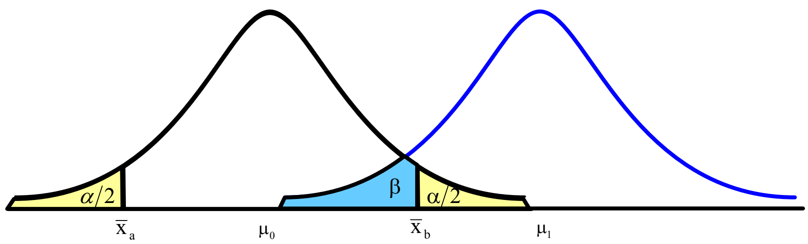


รูปที่ 1 แสดงบริเวณวิกฤต กรณีทดสอบ $H_0 : \mu = \mu_0$ เทียบกับ $H_1 : \mu \neq \mu_0$

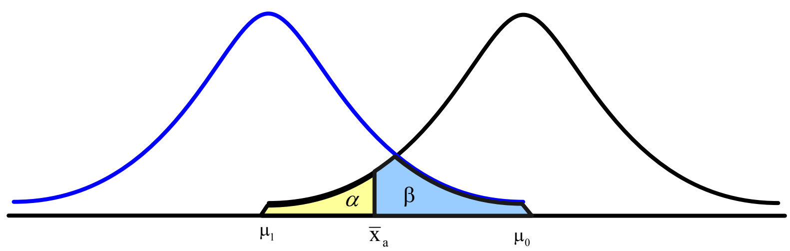
กรณีนี้เป็นการแสดงบริเวณวิกฤต ภายใต้สมมติฐาน H_0 เป็นจริง แต่ถ้า H_0 ไม่จริง ซึ่งอาจเป็นกรณีที่ $\mu_1 < \mu_0$ หรือ $\mu_1 > \mu_0$ ดังนั้นก็จะเกิดความคลาดเคลื่อนชนิดที่ 2 ขึ้น ซึ่งการหาความน่าจะเป็นของความคลาดเคลื่อนชนิดนี้พิจารณาจากบริเวณวิกฤตเดิมภายใต้เส้นโค้งการแจกแจงใหม่ที่เปลี่ยนพารามิเตอร์บอกตำแหน่ง (location parameter) ดังนี้



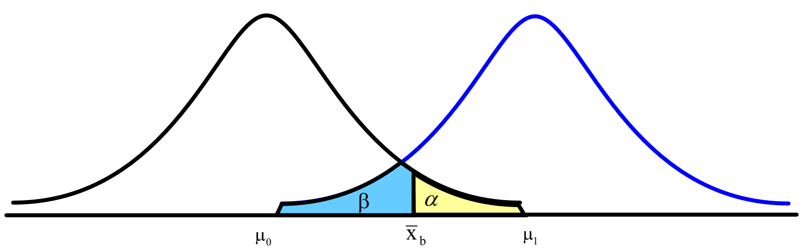
รูปที่ 2 แสดงความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 และ 2 ในกรณีทดสอบ $H_0 : \mu = \mu_0$ เทียบกับ $H_1 : \mu \neq \mu_0$ กรณี $\mu_1 < \mu_0$



รูปที่ 3 แสดงความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 และ 2 ในกรณีทดสอบ $H_0 : \mu = \mu_0$ เทียบกับ $H_1 : \mu \neq \mu_0$ กรณี $\mu_1 > \mu_0$

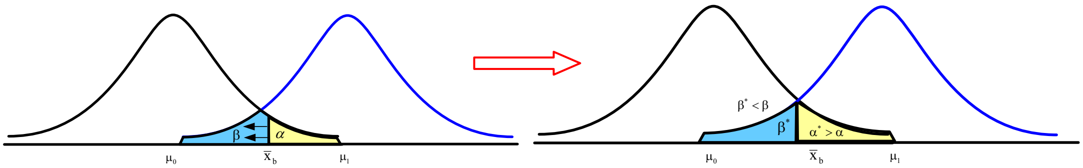


รูปที่ 4 แสดงความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 และ 2 ในกรณีทดสอบ $H_0 : \mu \geq \mu_0$ เทียบกับ $H_1 : \mu < \mu_0$ เมื่อ $\bar{x}_a = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$

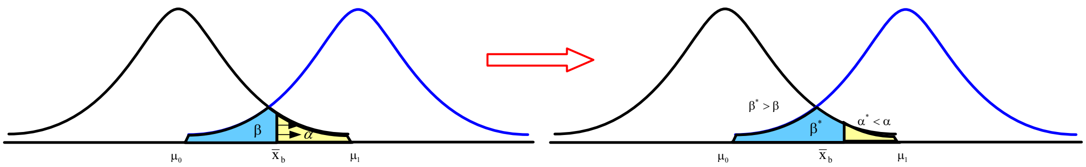


รูปที่ 5 แสดงความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 และ 2 ในกรณีทดสอบ $H_0 : \mu \leq \mu_0$ เทียบกับ $H_1 : \mu > \mu_0$ เมื่อ $\bar{x}_b = \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$

จากรูปที่ 5 ถ้าทำการเพิ่มค่า α ให้มากขึ้น จะพบว่าความน่าจะเป็น (พื้นที่) ของการปฏิเสธจะมากขึ้น ดังนั้นทำให้ β มีค่าลดลง ดังแสดงในรูปที่ 6 หรือถ้าลดค่า α ทำให้ β มีค่าเพิ่มขึ้น ดังแสดงในรูปที่ 7



รูปที่ 6 ความสัมพันธ์ระหว่าง α และ β กรณีทดสอบ $H_0 : \mu \le \mu_0$ เทียบกับ $H_1 : \mu > \mu_0$ เมื่อค่า α เพิ่มขึ้น



รูปที่ 7 ความสัมพันธ์ระหว่าง α และ β กรณีทดสอบ $H_0 : \mu \le \mu_0$ เทียบกับ $H_1 : \mu > \mu_0$ เมื่อค่า α ลดลง

5. การเปรียบเทียบการทดสอบ

ในการเปรียบเทียบการทดสอบจะพิจารณาจากฟังก์ชันกำลัง (Power Function) ซึ่งเป็นความน่าจะเป็นที่จะตัดสินใจปฏิเสธ H_0 นั่นคือ $P(\underset{\sim}{X} \in S_1)$ และขึ้นอยู่กับวงศ์ของการแจกแจง $\mathcal{P} = \{f(x,\theta), \theta \in \Omega\}$ หรือปริภูมิพารามิเตอร์ Ω ในกรณีที่สมมติฐานที่ทดสอบขึ้นอยู่กับ $\theta \in \Omega$ ฟังก์ชันกำลังของการทดสอบนิยมเขียนแทนด้วย $\beta(\theta) = P(\underset{\sim}{X} \in S_1 | \theta \in \Omega)$ และจะเรียกการทดสอบมีขนาด α (size of the test) สำหรับทดสอบ $H_0 : \theta \in \Omega_0$ ก็ต่อเมื่อ $\sup_{\theta \in \Omega_0} \beta(\theta) = \alpha$ และเรียกการทดสอบมีระดับ α (level of the test) ก็ต่อเมื่อ $\sup_{\theta \in \Omega_0} \beta(\theta) \le \alpha$ (Casella and Berger, 2001) ในการพิจารณาการทดสอบที่เหมาะสมจะพิจารณาจากกลุ่ม (class) ของการทดสอบเดียวกัน เช่น กลุ่มของการทดสอบที่มีขนาด α (แทนด้วย $\mathcal{T}_S$) หรือกลุ่มของการทดสอบระดับ α (แทนด้วย $\mathcal{T}_L$) (จะสังเกตเห็นได้ว่า $\mathcal{T}_S \subset \mathcal{T}_L$)

Lehmann (1986) และ Mukhopadhyay (2000) ได้กำหนดการทดสอบที่ดีที่สุด (the best test) สำหรับทดสอบ H_0 เทียบกับ H_1 จะพิจารณาเพียงกลุ่มของการทดสอบ $\mathcal{T}_L$ โดยที่สถิติทดสอบ $T \in \mathcal{T}_L$ จะเรียกว่าเป็นการทดสอบที่มีกำลังสูงสุดเสมอต้นเสมอปลายที่สุด หรือ uniformly most powerful test: UMP-Test ระดับ α ก็ต่อเมื่อ $\beta_T(\theta) \ge \beta_{T^*}(\theta)$ ทุก ๆ $\theta \in \Omega_1$ เมื่อ $T, T^* \in \mathcal{T}_L$ และถ้าการทดสอบ H_0 เทียบกับ H_1 เป็นการทดสอบเชิงเดี่ยว (simple hypothesis) จะเรียกสถิติทดสอบ T ที่เป็นสถิติทดสอบที่มีกำลังสูงสุดว่า the most powerful test: MP-Test ระดับ α

ประชุม (2545) นิยามบริเวณวิกฤตที่ดีที่สุด (best critical region: BCR) บนกลุ่มของการทดสอบ $\mathcal{T}_S$ ทำนองเดียวกันกับ Mukhopadhyay ในการทดสอบเชิงเดี่ยว H_0 เทียบกับ H_1 จะเรียกสถิติทดสอบ T ที่เป็นสถิติทดสอบที่มีกำลังสูงสุดว่า the most powerful test: MP α Test ขนาด α

ในการเปรียบเทียบสถิติทดสอบขึ้นอยู่กับว่าพิจารณาการทดสอบในกลุ่ม $\mathcal{T}_S$ หรือ $\mathcal{T}_L$ ซึ่งถ้าใช้ผลจากการจำลองข้อมูลโดยที่ปริภูมิตัวอย่างของผลลัพธ์ที่จะเกิดขึ้นแต่ละครั้ง คือ {ยอมรับ H_0 , ปฏิเสธ H_0 } โดยที่ถ้าทำการจำลอง K ครั้ง และกำหนดให้ $Y_i = 0$ ถ้าการทดสอบครั้งที่ i ยอมรับ H_0 หรือ $Y_i = 1$ ถ้าการทดสอบครั้งที่ i

ปฏิเสธ H_0 ดังนั้น $T_Y = \sum_{i=1}^{K} Y_i$ มีการแจกแจงทวินาม (K,p) เมื่อ $P(Y_i = 1) = p, \forall i = 1,2,\ldots,K$ และถ้าทำการจำลองข้อมูลภายใต้ H_0 เป็นจริง ในการทดสอบที่ระดับนัยสำคัญ α จะได้ว่า $P(Y_i = 1) = \alpha$ และได้ค่าประมาณของ α แทนด้วย $\hat{\alpha} = T_Y / K$ ทำนองเดียวกันถ้าทำการจำลองข้อมูลภายใต้ H_0 ไม่เป็นจริง ในการทดสอบที่ระดับนัยสำคัญ α เดียวกัน จะได้ว่า $P(Y_i = 1) = 1-\beta$ ซึ่งก็คือกำลังของการทดสอบ และได้ค่าประมาณของ $1-\beta$ แทนด้วย $1-\hat{\beta} = T_Y / K$

เกณฑ์ในการพิจารณาเปรียบเทียบการทดสอบจากการจำลองข้อมูล ส่วนมากผู้ศึกษาจะใช้ค่าความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 เป็นเกณฑ์ขั้นต้นในการจัดกลุ่มการทดสอบ แล้วจึงพิจารณาจากค่ากำลังของการทดสอบ ในการเลือกสถิติทดสอบที่เหมาะสมภายใต้สถานการณ์ในกลุ่มการทดสอบเดียวกัน ซึ่งมีผู้นิยมใช้หลายรูปแบบ อาทิเช่น บุรินทร์ (2549) Gu and Lee (2010) พิจารณาค่า $\hat{\alpha}$ ถ้าค่าที่ได้ใกล้เคียงกับขนาดของการทดสอบแสดงว่าวิธีการทดสอบนั้นมีประสิทธิภาพ

ดวงฤดี (2545) ชวนี และบุญอ้อม (2552) นิภาดา (2553) ปรีชา (2554) ทำการศึกษาและทดสอบค่า α โดยตั้งสมมติฐาน $H_0 : \alpha = \alpha_0$ เทียบกับ $H_1 : \alpha \neq \alpha_0$ เมื่อ $\alpha_0 = 0.01$ ที่ระดับนัยสำคัญ 0.01 และสรุปผลการทดสอบโดยวิธีจำลองข้อมูล ซึ่งในแต่ละสถานการณ์ กำหนด $K = 1{,}000$ รอบ โดยประมาณการแจกแจงของสถิติทดสอบด้วยการแจกแจงปรกติ ทำให้ได้กลุ่มการทดสอบขนาด α_0 ที่ทำให้ $\hat{\alpha} \in [0.002, 0.018]$ ทำนองเดียวกันในกรณีที่ $\alpha_0 = 0.05$ ที่ระดับนัยสำคัญ 0.05 ใช้เกณฑ์ $\hat{\alpha} \in [0.036, 0.063]$

ศศิกานต์ (2545) มงคล (2550) ภัทราพร (2551) ขจิตา (2552) ทำการศึกษากลุ่มการทดสอบจากการจำลองข้อมูล โดยใช้เกณฑ์จัดกลุ่มการทดสอบจากการทดสอบ $H_0 : \alpha \leq \alpha_0$ เทียบกับ $H_1 : \alpha > \alpha_0$ ที่ระดับนัยสำคัญ α^* โดยประมาณการแจกแจงของสถิติทดสอบด้วยการแจกแจงปรกติ ทำให้ได้กลุ่มการทดสอบระดับ α_0 จากนั้นพิจารณาสถิติทดสอบภายในกลุ่มดังกล่าว หากการทดสอบใดที่ให้อำนาจการทดสอบมากกว่าก็จะถือว่าการทดสอบนั้นเหมาะสม เช่น ถ้า $K = 1{,}000$ และ $\alpha_0 = 0.01, 0.05, 0.1$ ที่ระดับนัยสำคัญ 0.05 ใช้เกณฑ์ $\hat{\alpha} \leq 0.0152,\ 0.0613, 0.1156$ ตามลำดับ

สายทอง (2547) สถาพร (2547) นพดล (2553) ศิริรักษ์ (2553) Kim (2010) ใช้เกณฑ์ของ Bradley (1978, cited in Tomarken and Serlin, 1986) โดยที่พิจารณา $\hat{\alpha} \in [0.9\alpha_0, 1.1\alpha_0]$ หรือ $\hat{\alpha} \in [0.5\alpha_0, 1.5\alpha_0]$ และนิยมใช้ $\alpha_0 = 0.01,\ 0.05$ ใช้เกณฑ์ $\hat{\alpha} \in [0.005, 0.015],\ [0.025, 0.075]$ ตามลำดับ

สุพรรณี และวิชุดา (2540) พรรษณพร (2549) พลชาติ (2549) Vorapongsathorn et al. (2004) Sapchookul (2005) Kitidamrongsuk and Siripanich (2010) ใช้เกณฑ์ของ Cochran (1947) เมื่อ $\alpha_0 = 0.01,\ 0.05, 0.1$ ในการทดสอบสองทาง (two-sided test) ใช้เกณฑ์ $\hat{\alpha} \in [0.007, 0.015], [0.04, 0.06], [0.081, 0.119]$ ตามลำดับ ส่วนการทดสอบทางเดียว (one-sided test) จะใช้เกณฑ์ $\hat{\alpha}/2 \in [0.0025, 0.00075], [0.0175, 0.0325], [0.0425, 0.0575]$ เมื่อ $\alpha_0 = 0.01, 0.05,\ 0.1$ ตามลำดับ

Tomarken and Serlin (1986) ประยุกต์ใช้เกณฑ์ของ Cochran และ Bradley ในการกำหนดขอบเขตบนของ $\hat{\alpha}$ ที่ $\hat{\alpha} \in [0, 0.015], [0, 0.07]$ เมื่อ $\alpha_0 = 0.01, 0.05$ ตามลำดับ ซึ่งมีค่าใกล้เคียงกับการกำหนดเกณฑ์โดยใช้การทดสอบระดับ α_0

ในการนำผลการจำลองข้อมูลมาจัดกลุ่มการทดสอบว่าอยู่ใน $\mathscr{T}_S$ หรือไม่ จะตั้งสมมติฐานในการทดสอบ $H_0 : \alpha = \alpha_0$ เทียบกับ $H_1 : \alpha \neq \alpha_0$ และหากจัดกลุ่มการทดสอบว่าอยู่ใน $\mathscr{T}_L$ หรือไม่ จะตั้งสมมติฐาน $H_0 : \alpha \leq \alpha_0$ เทียบกับ $H_1 : \alpha > \alpha_0$ เมื่อกำหนดจำนวนรอบที่ทำซ้ำ K รอบ และกำหนดระดับนัยสำคัญ พบว่าภายใต้ H_0 เป็นจริง จะได้ $P(Y_i = 1) = \alpha$ และ $T_Y = \sum_{i=1}^{K} Y_i$ มีการแจกแจงทวินาม (K, α) ผู้ศึกษาเห็นว่าควรใช้สถิติทดสอบทวินามน่าจะเหมาะสม

เมื่อกำหนด $\alpha_0 = 0.01$ และ 0.05 และกำหนด $K = 200, 500, 1{,}000, 5{,}000$ และ $10{,}000$ จะได้ช่วงของค่า $\hat{\alpha}$ ดังแสดงในตารางที่ 1

ตารางที่ 1 ค่า $\hat{\alpha}$ ของการทดสอบที่อยู่ในกลุ่มการทดสอบ $\mathscr{T}_S$ และกลุ่มการทดสอบ $\mathscr{T}_L$ เมื่อจำแนกตามจำนวนรอบที่ทำซ้ำ K รอบ และระดับนัยสำคัญ จากการใช้การทดสอบทวินาม

α_0	K	$\hat{\alpha}\ (\mathscr{T}_S)$		$\hat{\alpha}\ (\mathscr{T}_L)$	
		ระดับนัยสำคัญ 0.01	**ระดับนัยสำคัญ 0.05**	**ระดับนัยสำคัญ 0.01**	**ระดับนัยสำคัญ 0.05**
0.01	200	[0.000, 0.03]	[0.000, 0.025]	[0.0, 0.03]	[0.0, 0.025]
	500	[0.002, 0.024]	[0.004, 0.022]	[0.0, 0.022]	[0.0, 0.018]
	1,000	[0.003, 0.018]	[0.005, 0.017]	[0.0, 0.018]	[0.0, 0.015]
	5,000	[0.0066, 0.0138]	[0.0074, 0.0128]	[0.0, 0.0134]	[0.0, 0.0124]
	10,000	[0.0076, 0.0127]	[0.0082, 0.0121]	[0.0, 0.0124]	[0.0, 0.0117]
0.05	200	[0.015, 0.09]	[0.015, 0.075]	[0.0, 0.09]	[0.0, 0.075]
	500	[0.028, 0.078]	[0.034, 0.072]	[0.0, 0.074]	[0.0, 0.066]
	1,000	[0.034, 0.069]	[0.037, 0.063]	[0.0, 0.067]	[0.0, 0.062]
	5,000	[0.042, 0.0578]	[0.0436, 0.0556]	[0.0, 0.0572]	[0.0, 0.0552]
	10,000	[0.0443, 0.0555]	[0.0459, 0.0544]	[0.0, 0.0551]	[0.0, 0.0536]

จากตารางที่ 1 จะพบว่าเมื่อระดับนัยสำคัญที่ทดสอบเพิ่มขึ้นจะทำให้ความกว้างช่วงของค่า $\hat{\alpha}$ ลดลง และถ้า α_0 เพิ่มขึ้นจะทำให้ความกว้างของช่วงเพิ่มมากขึ้น อีกทั้งถ้าจำนวนรอบการทำซ้ำเพิ่มมากขึ้นความกว้างช่วงของค่า $\hat{\alpha}$ จะลดลงเช่นกัน

6. สรุป

เมื่อเปรียบเทียบช่วงของค่า $\hat{\alpha}$ โดยใช้เกณฑ์ของ Cochran และ Bradley จะเห็นว่าทั้งสองเกณฑ์ไม่ได้พิจารณาจำนวนรอบของการทำซ้ำ ส่วนการพิจารณาจากการใช้สถิติทดสอบที่มีการแจกแจงทวินาม และการประมาณการทดสอบด้วยการแจกแจงปรกติ ช่วงของค่า $\hat{\alpha}$ มีค่าต่างกันไม่มากนักและช่วงที่ได้จะให้ความกว้างของ

ช่วงลดลง เมื่อจำนวนรอบการทำซ้ำมากขึ้น แต่เนื่องจาก $T_Y = \sum_{i=1}^{K} Y_i$ มีการแจกแจงทวินาม ผู้ศึกษาจึงเห็นว่าการใช้สถิติทดสอบ T_Y น่าจะเหมาะสมกับการจำลองข้อมูล

กิตติกรรมประกาศ

ผู้เขียนบทความขอขอบพระคุณผู้ประเมินบทความวิชาการที่กรุณาให้ข้อเสนอแนะในการปรับปรุงแก้ไขให้บทความนี้มีความสมบูรณ์ยิ่งขึ้น

เอกสารอ้างอิง

ขจิตา มัชฌิมา. (2552). ความเหมาะสมในการใช้สถิติทดสอบไคกำลังสองในการทดสอบความเป็นอิสระสำหรับข้อมูลจากการสำรวจตัวอย่าง. วารสารวิทยาศาสตร์มหาวิทยาลัยราชภัฏอุบลราชธานี 37(1): 151-152.

ชวนี สุภิรัตน์ และบุญอ้อม โฉมที. (2552). ประสิทธิภาพของสถิติทดสอบสำหรับแผนแบบแฟคทอเรียล กรณีที่มีตัวแปรร่วม. วารสารวิทยาศาสตร์มหาวิทยาลัยขอนแก่น 37(1): 106.

ดวงฤดี เหง้าพรมนิล. (2545). การวิเคราะห์ความแปรปรวนแบบปัจจัยเดียวเมื่อมีค่าสูญหายในแผนการทดลองแบบบล็อกสุ่มสมบูรณ์. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยขอนแก่น. ขอนแก่น: 72 หน้า.

นิภาดา จรัสเอี่ยม. (2553). การเปรียบเทียบสถิติทดสอบสำหรับตัวแบบการถดถอยปัวซงวางนัยทั่วไปที่มีศูนย์มากกับตัวแบบการถดถอยปัวซงวางนัยทั่วไปในกรณีที่มีข้อมูลผิดปกติในส่วนของค่าเฉลี่ย. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์. กรุงเทพฯ: 93 หน้า.

นพดล วันชนะชัย. (2553). การเปรียบเทียบตัวสถิติทดสอบไม่อิงพารามิเตอร์ระหว่างประชากร 2 กลุ่มเมื่อความแปรปรวนไม่เท่ากัน. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ: 71 หน้า.

ธีระพร วีระถาวร. (2536). การอนุมานเชิงสถิติขั้นกลาง: โครงสร้างและความหมาย. พิมพ์ครั้งที่ 2. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย. หน้า 172-175.

บุรินทร์ ราษฎร์อาศัย. (2549). สถิติทดสอบที่ใช้ในวิธีเจเนอรัลไรซ์พีวาลู สำหรับการทดสอบการเท่ากันของค่าเฉลี่ยจากการแจกแจงล็อกนอร์มอล 2 ประชากรที่อิสระกัน. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ. กรุงเทพฯ: 69 หน้า.

ประชุม สุวัตถี. (2545). ทฤษฎีการอนุมานเชิงสถิติ**.** พิมพ์ครั้งที่ 2**.** กรุงเทพฯ: โครงการส่งเสริมเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์. หน้า 275.

ปรีชา เครือโสม. (2554). การจำลองแบบเพื่อเปรียบเทียบประสิทธิภาพของการวิเคราะห์ความแปรปรวนสำหรับข้อมูลจำแนกทางเดียวเทียบกับการวิเคราะห์ความแปรปรวนของข้อมูลจำแนกสองทางในแผนแบบการทดลองแบบบล็อกสุ่มสมบูรณ์. Veridian E-Journal 4: 5-7.

พรรษณพร จันทร์ดี. (2549). สถิติทดสอบที่มีความแกร่งสำหรับการทดสอบความเท่ากันของความแปรปรวนของประชากร. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยเกษตรศาสตร์. กรุงเทพฯ: 89 หน้า.

พลชาติ หาญอนุรักษ์. (2549). การเปรียบเทียบตัวสถิติทดสอบเวลช์ บราวน์-ฟอร์สิติ และมาราสควิลโล สำหรับทดสอบความเท่ากันของค่าเฉลี่ย เมื่อความแปรปรวนของประชากรไม่เท่ากัน. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต,มหาวิทยาลัยเกษตรศาสตร์. กรุงเทพฯ: 77 หน้า.

มงคล ลีลาไพบูลย์. (2550). ตัวสถิติทดสอบมอนติคาร์โลอัตราส่วนภาวะน่าจะเป็นสำหรับแผนแบบจัตุรัสลาตินที่ปัจจัยทดลอง และปัจจัยบล็อกเป็นปัจจัยคงที่. วิทยานิพนธ์สถิติศาสตรมหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ: 159 หน้า.

ภัทราพร ทองนิ่ม. (2551). การเปรียบเทียบตัวสถิติทดสอบเอฟและตัวสถิติทดสอบมอนติคาร์โลอัตราส่วนความควรจะเป็นสำหรับแผนการทดลองแบบสุ่มตลอดที่ปัจจัยทดลองสุ่ม. วิทยานิพนธ์สถิติศาสตรมหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ: 130 หน้า.

สายทอง แจ่มใส. (2547). การเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบในการทดสอบภาวะสารูปสนิทดี. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยศิลปากร. กรุงเทพฯ: 129 หน้า.

สถาพร ชัยบุตร. (2547). อำนาจการทดสอบของตัวสถิติทดสอบในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเมื่อมีความแปรปรวนของความคลาดเคลื่อนไม่คงที่. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยเชียงใหม่. เชียงใหม่: 109 หน้า.

สุพรรณี อึ้งปัญสัตวงศ์. (2541). แนวคิดการเปรียบเทียบสถิติทดสอบโดยใช้การจำลองแบบปัญหาด้วยเทคนิคมอนติคาร์โล. วารสารวิทยาศาสตร์มหาวิทยาลัยขอนแก่น 26(1): 7-11.

สุพรรณี อึ้งปัญสัตวงศ์ และวิชุดา ไชยศิวามงคล. (2540). รายงานการวิจัยเรื่องผลกระทบของการปัดเศษของข้อมูลที่มีต่อระดับนัยสำคัญของสถิติทดสอบปกติบางตัว. มหาวิทยาลัยขอนแก่น: หน้า 22-23.

ศิริรักษ์ ไชยดี. (2553). การเปรียบเทียบสถิติทดสอบในการตรวจสอบความเหมาะสมของตัวแบบสมการถดถอยเชิงเส้น กรณีข้อมูลไม่มีการทำซ้ำ. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์. กรุงเทพฯ: 124 หน้า.

ศศิกานต์ คู่วัฒนา. (2545). การเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบบาร์ตเลต เลยาร์ดไคสแควร์และสแควร์แรงค์ สำหรับการทดสอบความเท่ากันของความแปรปรวนของประชากร. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ. กรุงเทพฯ: 242 หน้า.

Casella, G. and Berger, R.L. (2002). Statistics inference. Second Edition. Pacific Grove. CA. Duxbury. pp. 385.

Lehmann, E.L. (1986). Testing statistical hypotheses. Second Edition. New York: John Wiley. pp. 68-79.

Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics. 3: pp. 22-38.

Gu, X. and Lee, J.J. (2010). A simulation study for comparing testing statistics in response-adaptive randomization. BMC Medical Research Methodology 10: 1-14.

Kim, J. (2010). Controlling type 1 error rate in evaluating differential item functioning for four DIF methods: use of three procedures for adjustment of multiple item testing. Dissertation. Georgia State University. Georgia. Paper 140.

Kitidamrongsuk, P. and Siripanich, P. (2010). Selecting among families of lifetime distributions. In: Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA2010). Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia. pp. 367-376.

Mukhopadhyay, N. (2000). Probability and statistical inference. New York: Marcel Dekker. pp. 399-402.

Rizzo, M.L. (2008). Statistical computing with R. New York: Chapman & Hall/CRC. pp. 153-179.

Sapchookul, P. (2005). A comparison of tests for homogeneity of the risk difference when data are sparse. Master of Science (Biostatistics), Mahidol University. Bangkok: pp. 180.

Tomarken, A.J. and Serlin, R.C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychological Bulletin 99: 90-99.

Vorapongsathorn, T., Taejaroenkul, S. and Viwatwongkasem, C. (2004). Comparison of type I error under violation of assumptions. Songklanakarin Journal of Science and Technology 26: 537-547.

❑❑❑❑❑