

การประมาณค่าพารามิเตอร์ในการทดสอบสมมติฐาน สำหรับพารามิเตอร์ของการแจกแจงแบบเบร์นูลลี
โดยวิธีบูทสตรป

P – value Estimation in Hypothesis Testing for Parameter of Bernoulli Distribution
by Bootstrap Approach

พัชร มณีรัตน์ (Patcharee Maneerat)* ดร.ยงยุทธ ไชยพงศ์ (Dr.YoungYuth Chaiyapong)**

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อนำเสนอการประมาณค่าความน่าจะเป็นที่ได้จากชุดข้อมูลตัวอย่าง ว่ามีค่ามากหรือน้อยกว่าค่าที่กำหนด ภายใต้สมมติฐานหลัก H_0 ที่เป็นจริง (p – value) สำหรับพารามิเตอร์ของการแจกแจงแบบเบร์นูลลีด้วยวิธีบูทสตรป จากนั้นนำไปจัดทำข้อสรุปในการทดสอบสมมติฐาน สำหรับค่าพารามิเตอร์ในแต่ละสถานการณ์ ซึ่งจำแนกตามขนาดตัวอย่าง (n) ทั้งนี้เกณฑ์ที่ใช้ในการตรวจสอบความผิดพลาดและความเชื่อถือได้ดังกล่าว พิจารณาว่า ณ ระดับนัยสำคัญ α : ค่าประมาณ p – value ด้วยวิธีบูทสตรป ไม่มีโอกาสเกิดความผิดพลาดชนิดที่ 1 การศึกษาดำเนินการโดยใช้เทคนิคการจำลองข้อมูลที่มีการแจกแจงแบบเบร์นูลลี และทำการสุ่มตัวอย่างซ้ำแบบใส่คืน ด้วยโปรแกรมสำเร็จรูปทางสถิติ (MINITAB 14.0) ด้วยการซ้ำจำนวน 1,000 รอบ ในแต่ละสถานการณ์ ผลการศึกษาพบว่า ณ ระดับนัยสำคัญ 0.01 และ 0.05 ในการทดสอบสมมติฐาน $H_0 : P \leq P_0$ ให้ค่า p – value ที่ประมาณด้วยวิธีบูทสตรป ซึ่งไม่ได้อาศัยการลู่เข้า (Asymptotic) แสดงให้เห็นในทุกกรณีว่า ค่า p – value มากกว่าระดับนัยสำคัญที่กำหนด นั้นหมายความว่าไม่มีโอกาสที่เกิดความผิดพลาดชนิดที่ 1 และถือว่าไม่เป็นปฏิเสธสมมติฐานหลักไว้ก่อน ทั้งที่สมมติฐานหลักนั้นเป็นจริง

ABSTRACT

The purpose of this study is to propose on estimated p – value in hypothesis testing for parameter of Bernoulli Distribution with a bootstrap approach and inference in hypothesis testing for parameter in each situation with sample size (n). The p – value is used for detecting the errors and reliability. It is estimated with bootstrap approach that type I error is not happened. The simulation study was conducted by using MINITAB 14.0 program for generate the bernoulli random variables and repeated sampling with replacement by duplicated 1,000 times for each situation. The results of study shows that the p – value received from $H_0 : P \leq P_0$ at the 0.01 and 0.05 level of significant that is estimated with bootstrap approach which is not asymptotic are more than significant level in all cases. It means that type I error is not occurred and rejection H_0 is not expect before, even though a null hypotheses is true.

คำสำคัญ: วิธีบูทสตรป การลู่เข้า ความผิดพลาดชนิดที่ 1

Key Words: Bootstrap approach, Asymptotic, Type I error

* มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี สาขาวิชาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

** อาจารย์ประจำภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

บทนำ

การอนุมานเชิงสถิติสำหรับประชากรนั้นได้เป่าประสงค์หลัก คือการทราบค่าพารามิเตอร์ที่เป็นตัวกำกับพฤติกรรมของรูปแบบการแจกแจงความน่าจะเป็นที่กำกับการเกิดขึ้นของหน่วยประชากรเพื่อจัดทำข้อสรุปเกี่ยวกับพารามิเตอร์ดังกล่าว โดยใช้ระเบียบวิธีทางสถิติที่สำคัญ อันประกอบด้วย การประมาณค่าแบบจุด (Point Estimation) การประมาณค่าแบบช่วง (Interval Estimation) และการทดสอบสมมติฐาน (Hypothesis Testing) ขงยุทธ (2553) กล่าวว่า สำหรับการทดสอบสมมติฐาน เป็นการนำข้อมูลตัวอย่างจำนวนหนึ่ง ที่เก็บรวบรวมมาสร้างเป็นหลักฐานในรูปของสถิติทดสอบ เพื่อตัดสินชี้ชัดว่า ค่าพารามิเตอร์เป็นไปตามที่กำหนดไว้จริงหรือไม่ ซึ่งมีขั้นตอนที่สำคัญ คือ การตั้งสมมติฐานเกี่ยวกับพารามิเตอร์ แล้วกำหนดระดับนัยสำคัญของการทดสอบ (ความน่าจะเป็นในการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง) เพื่อใช้เป็นเกณฑ์ในการตัดสินใจ และเลือกใช้ตัวสถิติทดสอบที่เหมาะสม จากนั้นนำข้อมูลดังกล่าว มาทำการคำนวณหาค่าสถิติทดสอบ และค่าความน่าจะเป็นที่ค่าสถิติทดสอบมีค่ามาก หรือน้อยกว่า ค่าสถิติทดสอบที่คำนวณได้จากข้อมูล เรียกว่า “P – value” (Probability value) อีกนัยหนึ่งคือ ค่าความน่าจะเป็นที่ได้ค่าจากชุดของข้อมูลตัวอย่าง มีค่ามากหรือน้อยกว่าค่าที่กำหนดภายใต้สมมติฐานหลักที่เป็นจริง เพื่อนำไปสู่ข้อสรุป ในสมมติฐานเกี่ยวกับพารามิเตอร์ที่ตั้งขึ้น การดำเนินงานในทุกขั้นตอน จำเป็นต้องมีความเข้าใจใน นิยามและการจำแนกประเภทประชากร การกำหนดตัวแปรรูปแบบความน่าจะเป็นและพารามิเตอร์ที่กำกับรูปแบบการแจกแจง การตั้งสมมติฐานทางสถิติที่สอดคล้อง การเลือกใช้ตัวสถิติทดสอบที่เหมาะสม นำไปสู่การประเมินความเชื่อถือได้ของข้อสรุป โดยพิจารณาจากความผิดพลาดที่เกิดจากการทดสอบสมมติฐาน คือ ความผิดพลาดประเภทที่ 1 (หลักฐานจากตัวอย่างสุ่ม ทำให้มีการปฏิเสธสมมติหลัก ทั้งที่เป็นจริง หรือระดับ

นัยสำคัญ แทนสัญลักษณ์ด้วย α) และความผิดพลาดประเภทที่ 2 (หลักฐานจากตัวอย่างสุ่ม ทำให้มีการยอมรับสมมติหลัก ทั้งที่ไม่เป็นจริง แทนสัญลักษณ์ด้วย β) ซึ่งการวัดความผิดพลาดทั้ง 2 ประเภทดังกล่าว อยู่ในรูปของความน่าจะเป็นของการเกิดขึ้นของ 2 เหตุการณ์ แล้วพิจารณาพร้อมกับ อำนาจในการทดสอบ (หลักฐานจากตัวอย่างสุ่ม ทำให้มีการปฏิเสธสมมติหลัก ทั้งที่ไม่เป็นจริง แทนสัญลักษณ์ด้วย $1 - \beta$) ซึ่งใช้เป็นเกณฑ์ในการคัดเลือกตัวสถิติทดสอบ

สำหรับการศึกษาค้างนี้ ประชากรที่อยู่ภายใต้กรอบความสนใจ คือ ประชากรที่อยู่ภายใต้รูปแบบการแจกแจงแบร์นูลลี โดยมีพารามิเตอร์ P เป็นตัวกำกับพฤติกรรมของรูปแบบการแจกแจงดังกล่าว แสดงดังนี้ สำหรับ $X_1, X_2, \dots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันและมีรูปแบบการแจกแจงแบร์นูลลี $X_i \sim \text{Ber}(P)$ โดยที่ $i = 1, 2, \dots, n$ หากการทดลองสุ่มแต่ละครั้งเป็นอิสระต่อกัน และมีพารามิเตอร์ P เดียวกัน สามารถสร้างตัวประมาณแบบจุดสำหรับพารามิเตอร์ P ด้วยวิธี Maximum Likelihood Estimation : MLE คือ $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$ ซึ่งการจัดทำข้อสรุปในการศึกษาค้างนี้ใช้ระเบียบวิธีทางสถิติ คือ การทดสอบสมมติฐาน ซึ่งโดยทั่วไปวิธีในการสร้างตัวสถิติทดสอบมีหลายวิธี ด้วยแนวคิดที่แตกต่างกัน ปัจจุบันมีวิธีการพื้นฐาน 4 วิธีที่เหมาะสมกับสถานการณ์ที่แตกต่างกัน ได้แก่ วิธีอัตราส่วนควรจะเป็น (Likelihood Ratio Test : LRT), วิธีการสถิติทดสอบที่ไม่แปรเปลี่ยน (Invariant Test), วิธีการแบบเบย์เซียน (Bayesian Test) และวิธี Union – Intersection and Intersection – Union Test แต่วิธีการเบื้องต้นที่นิยมใช้กันอย่างแพร่หลาย ได้แก่ วิธีอัตราส่วนควรจะเป็น พิจารณาพร้อมกับการใช้หลักการความไม่ผันแปร แล้วสร้างตัวสถิติทดสอบ จากการปรับตัวสถิติทดสอบเดิม สำหรับประเด็นปัญหาสำคัญในการทดสอบสมมติฐาน เกี่ยวกับพารามิเตอร์ P ซึ่งตัวสถิติทดสอบถูกสร้างจากวิธีอัตราส่วนควรจะเป็น และอาศัยผลจากทฤษฎีบทการวัดแนวโน้มเข้าสู่ส่วนกลาง

(Central Limit Theorem : CLT) มาประยุกต์ใช้ เมื่อสมมติฐานที่ต้องการทดสอบ คือ $H_0 : P = P_0$ vs. $H_1 : P \neq P_0$ และตัวสถิติทดสอบที่ใช้กันในทางปฏิบัติคือ Z - Test โดยที่ $Z = \frac{\hat{p} - P_0}{\sqrt{\hat{p}(1-\hat{p})/n}}$ ซึ่งค่าความแปรปรวนได้จากการใช้ตัว

ประมาณแบบ MLE สำหรับพารามิเตอร์ P และใช้ผลจากทฤษฎีแวนโน้มเข้าสู่ส่วนกลาง เพื่อประมาณรูปแบบการแจกแจงของตัวสถิติทดสอบดังกล่าว เพื่อให้สะดวกต่อการใช้ประโยชน์ สำหรับ Chi - square Test เป็นผลสืบเนื่องจากการใช้ Z - Test ส่งผลให้ตัวสถิติทดสอบทั้งสองมีรูปแบบการแจกแจงปกติมาตรฐานโดยประมาณ เมื่อตัวอย่างมีขนาดใหญ่ ($n \rightarrow \infty$) ทำให้การสรุปเกี่ยวกับอาณาเขตวิกฤต และค่า p - value มีค่าไม่ถูกต้อง ในส่วน Score Test ตัวสถิติทดสอบคือ $Z = \frac{\hat{p} - P_0}{\sqrt{P_0(1-P_0)/n}}$ ซึ่งค่าความแปรปรวน

ได้มาจากพารามิเตอร์ที่อิงจากสมมติฐานหลัก H_0 และมีการแจกแจงปกติมาตรฐานโดยประมาณเช่นกัน ผลลัพธ์คือ เกิดการปฏิเสธสมมติฐาน H_0 ไว้ก่อน และสิ่งที่ต้องคำนึงคือ ขนาดตัวอย่างต้องมากพอ มิฉะนั้นจะส่งผลให้เกิดความผิดพลาดประเภทที่ 1 สูงมาก เนื่องจากการใช้ผลจากทฤษฎีการวัดแวนโน้มเข้าสู่ส่วนกลาง เพื่อประมาณรูปแบบการแจกแจงของตัวสถิติทดสอบดังกล่าว นอกจากนั้นแล้วค่าพารามิเตอร์ P ต้องมีค่าเข้าใกล้ 0.5 จึงจะทำให้การแจกแจงทวินามมีลักษณะสมมาตรสู่เข้าสู่การแจกแจงปกติมาตรฐาน เมื่อ $n \rightarrow \infty$ ประกอบกับการใช้ข้อมูลตัวอย่างจำนวนหนึ่งไปจัดทำข้อสรุปเกี่ยวกับประชากรอนันต์ อีกแนวทางหนึ่งเพื่อหลีกเลี่ยงการใช้รูปแบบการแจกแจงโดยประมาณ ซึ่ง Hutson (2006) พิจารณาการใช้ Exact Test ในการหาค่า p - value ซึ่งอยู่ในรูปของความน่าจะเป็นของส่วนปลายหาง จากค่าของตัวแปรสุ่ม $Y = \sum_{i=1}^n X_i$ ซึ่งมีการแจกแจงทวินาม ผลลัพธ์คือ เกิดการปฏิเสธสมมติฐาน H_0 ไว้ก่อน หมายความว่า ณ

ระดับนัยสำคัญ α หากต้องการตรวจสอบ $\Pr(Y \geq y) = \alpha$ ซึ่งไม่มีโอกาสที่จะให้ความน่าจะเป็นสะสมเท่ากับ α ได้ ซึ่งในทางปฏิบัติเป็นการใช้ข้อมูลตัวอย่างเพียงจำนวนหนึ่ง หากตัวอย่างขนาดเล็ก ค่าตัวแปรสุ่มที่เป็นไปได้มีเพียงไม่กี่ค่าเท่านั้น ซึ่งไม่มีค่าของตัวแปรสุ่ม Y ในปริภูมิตัวอย่างที่ให้ความน่าจะเป็นสะสมได้เท่ากับ α ได้ เนื่องจาก Y เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง (Discrete Random Variable)

จากศึกษาของ Agresti & coull (1998) พบว่าวิธีการ Exact Test ให้ผลการทดสอบสมมติฐานที่ยังคงมีการปฏิเสธสมมติฐานหลักไว้ก่อน เนื่องจาก X_i เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง ภายใต้การแจกแจงแบร์นูลลี Brown (2001) พบว่าช่วงของ Clopper - Pearson (Exact) ไม่เกิดการปฏิเสธสมมติฐานหลักไว้ก่อน แต่ไม่เหมาะสมในทางปฏิบัติ เนื่องจากกำหนดให้ $C(\pi, n) \geq 1 - \alpha$ เมื่อ $C(\pi, n)$ คือ ความน่าจะเป็นในการครอบคลุม (Coverage Probability) ค่าพารามิเตอร์ของการประมาณค่าแบบช่วง สำหรับพารามิเตอร์ π สำหรับ Hutson (2006) พบว่า การศึกษาของ Neyman (1935) มีความสอดคล้องกับ Agresti & coull (1998) และเชื่อว่าเป็นวิธีการที่เหมาะสมในทางปฏิบัติ สำหรับการใช้ Exact Test สำหรับ Lloyd (2008) กล่าวว่า การประมาณค่า p - value ด้วยวิธีนูนแสดงรูปให้ความถูกต้องแม่นยำกว่า Standard p - value ถึงแม้ว่าตัวอย่างมีขนาดเล็ก จึงเป็นเหตุผลที่วิธีนูนแสดงรูปเหมาะสมสำหรับพิจารณาพร้อมกับหลักการของการถ่วงเข้า Lloyd (2010) พบว่าการประมาณค่า p - value ด้วยวิธีนูนแสดงรูปมีคุณสมบัติเหมาะสมสำหรับการใช้ความถี่ (Frequentist properties) สามารถลดความคลาดเคลื่อนได้ ซึ่งแตกต่างจากค่า p - value จากการใช้ตัวสถิติทดสอบอื่น เช่น Likelihood root, Score และ Wald statistics

ดังนั้นการศึกษาค้นคว้าครั้งนี้มีจุดประสงค์เพื่อนำเสนอเกี่ยวกับการทดสอบสมมติฐาน สำหรับพารามิเตอร์ P ของการแจกแจงแบร์นูลลี โดยการหาค่าประมาณ p - value ด้วยวิธีนูนแสดงรูป ผ่านตัวสถิติ

$T = \sum_{i=1}^n X_i$ ซึ่งอยู่ภายใต้การแจกแจงทวินาม เนื่องจาก
หลีกเลี่ยงปัญหาอันอันเกิดจาก การใช้ตัวสถิติทดสอบที่
ต้องทราบรูปแบบการแจกแจงที่แท้จริง ซึ่งจะส่งผล
กระทบต่อเกิดความผิดพลาดประเภทที่ 1

วัตถุประสงค์

1. เพื่อนำเสนอการประมาณค่า P- value
สำหรับพารามิเตอร์ของการแจกแจงแบร์นูลลี ด้วย
วิธีการบูทสเตรป
2. เพื่อจัดทำข้อสรุปในการทดสอบ
สมมติฐาน สำหรับพารามิเตอร์ ของรูปแบบการแจกแจง
แบร์นูลลีในแต่ละสถานการณ์ จำแนกตามระดับ
นัยสำคัญ (α) และขนาดตัวอย่าง (n)

การทดสอบสมมติฐาน (Hypothesis Testing)

การทดสอบสมมติฐานทางสถิติ ซึ่งเป็น
ระเบียบวิธีทางสถิติอีกประเภทหนึ่งในการจัดทำ
ข้อสรุปเกี่ยวกับประชากรนั้น นั่นคือแทนที่ใช้ข้อมูล
ตัวอย่างจำนวนหนึ่งทีเก็บรวบรวมมาทำการประมาณ
ค่าพารามิเตอร์ดังเช่นการประมาณค่าแบบจุด แต่จะเป็น
การนำข้อมูลดังกล่าวมาสร้างเป็นหลักฐานในรูป
ของสถิติทดสอบสำหรับตัดสินใจว่าค่าพารามิเตอร์ มี
ค่าเป็นตามที่กำหนดไว้จริงหรือไม่ หากมีการตั้ง
ข้อสังเกตเกี่ยวกับค่าพารามิเตอร์ θ ว่ามีค่าเป็นดังที่
สงสัยจริงหรือไม่ ยกตัวอย่างเช่นการตั้งข้อสังเกตว่า
กระบวนการแบร์นูลลีที่อยู่ภายใต้กรอบความสนใจมี
ค่าพารามิเตอร์ $P = 0.5$ จริงหรือไม่ การกระทำใน
ลักษณะเช่นนี้ในทางสถิติศาสตร์เรียกว่าการ
ตั้งสมมติฐานเกี่ยวกับค่าพารามิเตอร์ P ของรูปแบบ
ความน่าจะเป็นแบบแบร์นูลลีที่กำกับการเกิดขึ้นของ
ประชากรว่ามีค่าเป็น 0.5 ข้อสังเกตเกี่ยวกับพารามิเตอร์
ในลักษณะดังกล่าวข้างต้นนี้ในทางสถิติเรียกว่า
“สมมติฐานทางสถิติ”

หลักการในการทดสอบสมมติฐาน คือ เป็น
การตัดสินใจระหว่างสองสมมติฐานซึ่งแสดง

พารามิเตอร์ที่สนใจ : θ ตั้งแต่ 1 ตัวขึ้นไป ประกอบด้วย
สมมติฐานหลัก (Null Hypothesis) แทนด้วยสัญลักษณ์
 $H_0 : \theta \in \omega$ และสมมติฐานทางเลือก (Alternative
Hypothesis) แทนด้วยสัญลักษณ์ $H_1 : \theta \in \Omega - \omega$ โดยที่
 Ω เป็นเซตของพารามิเตอร์ θ ที่เป็นไปได้ทั้งหมด
เรียกว่า ปริภูมิพารามิเตอร์ (Parameter Space) และ ω
เป็นสับเซตของ Ω สำหรับการตัดสินใจเพื่อจะชี้ชัดว่า
พารามิเตอร์ที่สนใจเป็นไปตามสมมติฐาน H_0 หรือ H_1
นั้น โดยพิจารณาจากตัวแปรสุ่ม $X_1, X_2, \dots, X_n$ ภายใต้
รูปแบบการแจกแจงใดๆ $f(x; \theta)$ ซึ่งเป็นรูปแบบความ
น่าจะเป็นที่กำกับกลไกก่อนกำเนิดหน่วยประชากร
โดยทั่วไปกำหนดให้ C เป็นตัวสถิติ (Test Statistics
: $T(X)$) ที่มีค่ามาก เรียกว่าเป็นอาณาเขตวิกฤต (Critical
Region/Reject Region) และ $\bar{C}$ เป็นอาณาเขตในการ
ยอมรับสมมติฐาน H_0 เมื่อ $X_1, X_2, \dots, X_n$ ซึ่งเป็นสับ
เซตของ C แล้วถ้าหาก $X \in C$ จะทำการปฏิเสธ
สมมติฐาน H_0 (Garthwaite, 2002)

การประเมินความผิดพลาดและความเชื่อถือ ได้ของการทดสอบสมมติฐาน

ในการทดสอบสมมติฐานทางสถิติโดยนำ
ข้อมูลตัวอย่างมาผ่านกระบวนการที่กำหนดเพื่อชี้ชัดว่า
สมมติฐานหลักเป็นจริงหรือไม่นั้นย่อมมีความผิดพลาด
เกิดขึ้นอันเนื่องมาจากการใช้ข้อมูลตัวอย่างเพียงจำนวน
หนึ่งไปทำการจัดทำข้อสรุปเพื่ออธิบายสภาพความ
เป็นไปของประชากรนั้น Casella and Berger (2002)
กล่าวว่า สำหรับวิธีในการประเมินความผิดพลาดของ
การทดสอบสมมติฐานนั้น อยู่ในรูปแบบที่พิจารณาจาก

1. ความน่าจะเป็นของการเกิดความผิดพลาด
และฟังก์ชันการทดสอบ (Error Probability and Power
function)
2. การทดสอบที่ให้อำนาจการทดสอบสูงสุด
(Most Powerful Tests)
3. ขนาดตัวอย่างสำหรับการทดสอบวิธี
Union – Intersection และ Intersection - Union test
4. ค่า p – value

5. ฟังก์ชันการสูญเสียที่เหมาะสม (Loss Function Optimality)

สำหรับในการศึกษาครั้งนี้จะกล่าวถึงเฉพาะการประเมินความผิดพลาดของการทดสอบสมมติฐานที่พิจารณาจากค่าประมาณ p - value ส่งผลทำให้เกิดความผิดพลาดประเภทที่ 1 (α) หรือไม่

การประมาณค่า p -value ด้วยวิธีบูทสเตรป

เมื่อ $X_1, X_2, \dots, X_n$ เป็นตัวแปรสุ่มที่มีรูปแบบการแจกแจงใดๆ ซึ่ง θ เป็นพารามิเตอร์กำกับรูปแบบการแจกแจงนั้น ๆ หากต้องการทดสอบสมมติฐานสำหรับพารามิเตอร์ดังกล่าว แต่รูปแบบการแจกแจงนั้นไม่สามารถสร้างตัวสถิติทดสอบ หรือไม่สามารถหา รูปแบบการแจกแจงที่แท้จริงของตัวสถิติทดสอบนั้นได้ ในกรณีนี้ คือการใช้ค่าของตัวแปรสุ่ม ที่เกิดการสุ่มตัวอย่างซ้ำแบบใส่คืน ด้วยวิธีบูทสเตรป เพื่อใช้ในการตัดสินใจในการทดสอบสมมติฐานนั้น โดยพิจารณาจากค่า p -value ซึ่งเป็นค่าความน่าจะเป็นของการเกิดขึ้นของเหตุการณ์ที่ผิดแปลกไปจากเหตุการณ์ส่วนใหญ่ ภายใต้สมมติฐานหลักที่เป็นจริง นั่นคือ ณ ระดับนัยสำคัญ α : เมื่อตัวอย่างสุ่มนั้นเกิดค่าผิดปกติ (Outlier) ไปจนถึงค่าสุดโต่ง (Extremes Value) ภายใต้ข้อมูลที่เก็บรวบรวมมา

ค่า p -value คือค่าความน่าจะเป็นที่ได้จากชุดของข้อมูลตัวอย่างซึ่งมีค่ามาก หรือน้อยกว่าค่าที่กำหนด (Cut off point) ภายใต้สมมติฐานหลักที่เป็นจริง ยกตัวอย่างเช่น หากต้องการทดสอบสมมติฐาน $H_0 : P = 0.5$ หรือไม่ เมื่อมีข้อมูลตัวอย่างชุดหนึ่งนำมาหาค่าประมาณโดยใช้ตัวประมาณที่สร้างจากวิธี MLE คือ $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$ ได้ $\hat{p} = 0.49$ แล้วนำไปหาค่า p -value จากการใช้ตัวสถิติทดสอบใดๆ แล้วพบว่า p -value = 0.795 นั้นหมายความว่า ค่าประมาณ $\hat{p} = 0.49$ มีโอกาสเกิดขึ้นมากถึง 79.5% เมื่อค่าพารามิเตอร์เป็นไปตามสมมติฐาน $H_0 : P = 0.5$ นั่นคือ มีโอกาสที่จะเกิดขึ้นสูง หากสมมติฐานหลักเป็นจริง ดังนั้นจึงมีหลักฐานเพียงพอที่ยอมรับสมมติฐาน

ดังกล่าว จึงอาจจะกล่าวได้ว่าค่า p -value เป็นค่าความน่าจะเป็นที่แสดงถึงหลักฐานที่จะนำมาแข่งกับสมมติฐานหลัก ว่าเหตุการณ์นั้นมีโอกาสเกิดขึ้นได้มากน้อยเพียงไร ภายใต้สมมติฐานหลักนั้นเป็นจริง

การทดสอบสมมติฐาน ในกรณีที่การแจกแจงประกอบด้วยพารามิเตอร์เพียงตัวเดียว ด้วยวิธีบูทสเตรป

โดยทั่วไป การทดสอบสมมติฐาน เป็นการอนุมานทางสถิติสำหรับประชากรหนึ่งแล้วพิจารณา ร่วมกับวิธีบูทสเตรป จะกำหนดรูปแบบการแจกแจงของประชากรที่ต้องการศึกษา F_0 เพื่อต้องการสร้างชุดตัวอย่างแบบบูทสเตรป และตัวสถิติ T สำหรับพารามิเตอร์ที่ใช้ในการทดสอบสมมติฐานหลัก H_0 อยู่ภายใต้รูปแบบการแจกแจงของประชากรโดยประมาณ $\hat{F}_0$ ซึ่งไม่ใช่ฟังก์ชันรูปแบบการแจกแจงของข้อมูล (Empirical Distribution Function: $\hat{F}$) หากใช้พารามิเตอร์ภายใต้ $\hat{F}$ จะไม่มีโอกาสที่จะปฏิเสธสมมติฐานหลัก เนื่องจากทราบพารามิเตอร์ที่ชัดเจน ภายใต้ฟังก์ชันรูปแบบการแจกแจงของข้อมูลชุดนั้น การเลือกใช้ $\hat{F}_0$ สามารถกำหนดได้โดยตรง แต่ต้องอาศัยทักษะ และการประยุกต์ใช้อย่างเหมาะสม หาก $\hat{F}_0$ ไม่ทราบอย่างชัดเจน จะเพิ่มเกณฑ์บางอย่างเพื่อที่จะให้สอดคล้องกับข้อมูลตัวอย่างนั่นคือ การเพิ่มขนาดของชุดตัวอย่างแบบบูทสเตรป (B) ให้มากที่สุด ซึ่งแต่ละชุดตัวอย่างแบบบูทสเตรป ถูกทำการสุ่มซ้ำจาก $\hat{F}_0$ และทราบค่าสถิติในแต่ละชุด

กำหนดให้

t_i^* เป็นค่าสถิติ สำหรับชุดตัวอย่างแบบบูทสเตรป ชุดที่ i ; $i = 1, 2, \dots, B$

t เป็นค่าสถิติจากสถานการณ์ภายใต้สมมติฐานหลัก H_0

หากสมมติ t เป็นค่าสถิติมากที่สุดภายใต้พารามิเตอร์ที่สอดคล้องกับ H_0 ดังนั้นค่า p -value แสดงดังนี้

$$p = \frac{1 + (\text{Number of } t_i^* \geq t)}{B+1}$$

ดังนั้นจะปฏิเสธภายใต้สมมติฐานหลัก ณ ระดับนัยสำคัญ 100p% (Garthwaite, 2002)

การทดสอบสมมติฐานในกรณีที่มีการแจกแจงประกอบด้วย พารามิเตอร์รบกวน (Nuisance parameter) ด้วยวิธีบูทสเตรป

สมมติ $Y_1, Y_2, \dots, Y_n$ เป็นตัวแปรสุ่มที่มีรูปแบบการแจกแจง $\pi(y; \psi, \lambda)$ เมื่อ λ เป็นพารามิเตอร์รบกวน และต้องการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ $H_0: \psi = \psi_0$ โดยใช้ตัวสถิติ T สำคัญจากข้อมูลตัวอย่างให้ค่า $t = T(y)$ เป็น

$$S(t; \lambda) = \Pr(T(Y) \geq t; \psi_0, \lambda)$$

ในขณะที่ต้องการให้ $S(t; \hat{\lambda})$ อยู่เข้าสู่ $S(t; \lambda)$ ซึ่งประเด็นปัญหาที่เกี่ยวกับการหาค่า p-value ด้วยวิธีบูทสเตรป วิธีที่จะอยู่ได้ นั่นคือการอยู่ด้วยรูปแบบการแจกแจง Uniform. ซึ่ง Lee and Young (2005) กล่าวว่า ตัวสถิติจากวิธี Likelihood root และตัวสถิติใดๆ สามารถเกิดความผิดพลาดจากการที่ต้องการให้ $S(t; \hat{\lambda})$ อยู่เข้าสู่ $S(t; \lambda)$ เท่ากับ $O_p(m^{-1/2})$ ณ ระดับนัยสำคัญ α แสดงดังนี้

$$\Pr(S(T; \hat{\lambda}_0) \leq \alpha; \psi_0, \lambda) = \alpha + O(m^{-3/2})$$

เมื่อ $\alpha \in [0, 1]$ (Lloyd, 2010)

ขอบเขตและวิธีการวิจัย

ขอบเขตการวิจัย

1. กำหนดให้ $X_1, X_2, \dots, X_n$ เป็นตัวแปรสุ่มจากประชากร ภายใต้การแจกแจงเบอร์นูลลี ด้วยค่าพารามิเตอร์ P คือ $P = 0.1, 0.3, 0.5, 0.7, 0.9$ จากสมมติฐานที่ต้องการทดสอบคือ $H_0: P \leq P_0$

2. ขนาดตัวอย่าง (n) ที่ศึกษาแบ่งเป็น 3 กลุ่มคือ กลุ่มที่ 1 ขนาดตัวอย่าง $n \leq 30$ กลุ่มที่ 2 ขนาดตัวอย่าง $30 < n \leq 100$ กลุ่มที่ 3 ขนาดตัวอย่าง $n > 100$

3. ระดับนัยสำคัญ α เท่ากับ 0.01, 0.05

วิธีการวิจัย

1. จำลองข้อมูลตามสถานการณ์ที่กำหนดไว้ตามขอบเขตการวิจัย

2. จากนั้นคำนวณค่า p-value ด้วยวิธีบูทสเตรป คือ ทำการสุ่มตัวอย่างซ้ำแบบใส่คืน (With replacement) จากชุดตัวอย่างเดิม คือ X_{jb}^* โดยที่ $j = 1, 2, \dots, m$ $b = 1, 2, \dots, B$ ซึ่ง $B = 1,000$ ชุด และแต่ละชุดมีขนาด m โดยที่ $m = n$ ในสถานการณ์หนึ่ง จากนั้นหาค่าสถิติ t จากค่าพารามิเตอร์ตามสมมติฐานหลัก H_0 เพื่อกำหนดเป็นจุดตัด (Cut off point) ในการประมาณค่า p-value ด้วยวิธีบูทสเตรป

กำหนดให้

X_i แทนตัวแปรสุ่มภายใต้รูปแบบการแจกแจงเบอร์นูลลีลำดับที่ i โดยที่

$$i = 1, 2, \dots, n$$

X_{jb}^* แทนตัวแปรสุ่มภายใต้รูปแบบการแจกแจงเบอร์นูลลี จากวิธีบูทสเตรป ชุดที่ b ลำดับการสุ่มที่ j โดยที่

$$j = 1, 2, \dots, m \text{ และ } b = 1, 2, \dots, B$$

$$t = \sum_{i=1}^n X_i \text{ แทนผลรวมค่าของตัวแปรสุ่มจาก}$$

สมมติฐานหลัก H_0

$$t_b^* = \sum_{j=1}^m X_{jb}^* \text{ แทนผลรวมค่าของตัวแปรสุ่มข้างต้นโดย}$$

วิธีบูทสเตรปในชุดที่ b แสดงดังนี้

$$\begin{aligned} X_{11}^*, X_{21}^*, \dots, X_{m1}^* &\Rightarrow t_1^* = \sum_{j=1}^m X_{j1}^* \\ X_{12}^*, X_{22}^*, \dots, X_{m2}^* &\Rightarrow t_2^* = \sum_{j=1}^m X_{j2}^* \\ &\dots \\ X_{1B}^*, X_{2B}^*, \dots, X_{mB}^* &\Rightarrow t_B^* = \sum_{j=1}^m X_{jB}^* \end{aligned}$$

3. พิจารณา ค่า t_b^* และ t หาก $t_b^* \geq t$ ตัวแปรใหม่จะถูกกำหนดให้มีค่าเป็น “1” ถือว่าเป็นผลลัพธ์ที่ “success” ถ้าหาก $t_b^* < t$ ตัวแปรใหม่จะถูกกำหนดให้มีค่าเป็น “0” ถือว่าเป็นผลลัพธ์ที่ “failure”

4. นำค่าตัวแปรใหม่ทั้งสิ้น 1,000 ค่าจากจำนวนชุดตัวอย่างแบบสุ่มจำนวน 1,000 ชุด เพื่อนำไปหาค่า P- value จากวิธีbootstrap ได้ 1 ค่าจากฟังก์ชัน ดังนี้

$$\Pr(t_b^* \geq t) = \frac{1 + \text{Number}(t_b^* \geq t)}{B + 1}$$

5. กระทำซ้ำจากข้อ 1 - 4 จำนวน 1000 รอบ ได้ค่า p - value ทั้งสิ้น 1000 ค่า แล้วนำไปหาค่าเฉลี่ยเพื่อประมาณค่า p - value ด้วยวิธีbootstrap สำหรับสถานการณ์ นั้นๆ

6. ณ ระดับนัยสำคัญ α นำค่า p - value ด้วยวิธีbootstrap ไปเปรียบเทียบกับระดับนัยสำคัญที่กำหนด

7. สรุปผลการศึกษา

ผลการวิจัยและการอภิปรายผล

ตารางที่ 1 แสดงค่า P - value จากตัวสถิติ $T = \sum_{i=1}^n X_i$

ซึ่งประมาณด้วยวิธีbootstrap จากสมมติฐาน

$H_0 : P \leq P_0$ จำแนกตามขนาดตัวอย่าง

(n) ค่าพารามิเตอร์ (P)

n	P	ค่า p- value	n	P	ค่า p- value
10	0.1	0.497	60	0.1	0.524
	0.3	0.555		0.3	0.527
	0.5	0.599		0.5	0.534
	0.7	0.630		0.7	0.553
	0.9	0.694		0.9	0.600
20	0.1	0.543	70	0.1	0.521
	0.3	0.565		0.3	0.521
	0.5	0.554		0.5	0.522
	0.7	0.588		0.7	0.544
	0.9	0.654		0.9	0.602
30	0.1	0.515	80	0.1	0.515
	0.3	0.538		0.3	0.513
	0.5	0.549		0.5	0.539
	0.7	0.570		0.7	0.539
	0.9	0.650		0.9	0.575
40	0.1	0.510	90	0.1	0.522
	0.3	0.517		0.3	0.506
	0.5	0.558		0.5	0.532
	0.7	0.559		0.7	0.551
	0.9	0.630		0.9	0.572
50	0.1	0.520	100	0.1	0.502
	0.3	0.524		0.3	0.527
	0.5	0.549		0.5	0.518
	0.7	0.562		0.7	0.533
	0.9	0.591		0.9	0.567

ตารางที่ 1 แสดงค่า P – value จากตัวสถิติ $T = \sum_{i=1}^n X_i$

ซึ่งประมาณด้วยวิธีบูทสเตรป จากสมมติฐาน

$H_0 : P \leq P_0$ จำแนกตามขนาดตัวอย่าง

(n) ค่าพารามิเตอร์ (P) (ต่อ)

n	P	ค่า p- value	n	P	ค่า p- value
200	0.1	0.518	700	0.1	0.488
	0.3	0.520		0.3	0.518
	0.5	0.506		0.5	0.533
	0.7	0.528		0.7	0.523
	0.9	0.557		0.9	0.543
300	0.1	0.506	800	0.1	0.499
	0.3	0.515		0.3	0.505
	0.5	0.507		0.5	0.528
	0.7	0.532		0.7	0.515
	0.9	0.531		0.9	0.542
400	0.1	0.500	900	0.1	0.484
	0.3	0.526		0.3	0.500
	0.5	0.528		0.5	0.508
	0.7	0.536		0.7	0.497
	0.9	0.536		0.9	0.526
500	0.1	0.506	1000	0.1	0.514
	0.3	0.499		0.3	0.497
	0.5	0.521		0.5	0.523
	0.7	0.516		0.7	0.522
	0.9	0.534		0.9	0.517
600	0.1	0.511			
	0.3	0.518			
	0.5	0.494			
	0.7	0.534			
	0.9	0.536			

ผลการประมาณค่า p – value ด้วยวิธีบูทสเตรป จากสมมติฐาน $H_0 : P \leq P_0$ สำหรับพารามิเตอร์ P โดยใช้เกณฑ์การเปรียบเทียบกับระดับนัยสำคัญที่กำหนด พบว่า

เมื่อเทียบกับระดับนัยสำคัญ 0.01 และ 0.05; พบว่าหาก $P = 0.1, 0.3, 0.5, 0.7, 0.9$ ให้ค่า p – value ที่ถูกประมาณด้วยวิธีบูทสเตรป มากกว่าระดับนัยสำคัญดังกล่าว ซึ่งนำไปสู่การยอมรับสมมติฐานหลัก H_0 ในทุก ๆ ขนาดตัวอย่าง ซึ่งจะไม่เกิดเหตุการณ์ที่จะปฏิเสธสมมติฐานหลัก H_0 ทั้งที่สมมติฐานหลักนั้นเป็นจริง และขนาดตัวอย่าง $n=10$ เมื่อ $P=0.1$ และ $P=0.9$ ค่าประมาณที่ได้จากข้อมูลตัวอย่างมีโอกาสที่เกิดขึ้น เท่ากับ 49.7% และ 69.4 % ภายใต้สมมติฐานหลักที่เป็นจริง ซึ่งเป็นโอกาสที่เกิดขึ้นได้สูง จึงมีหลักฐานเพียงพอที่ยอมรับสมมติฐานนั้น แม้จะมีตัวอย่างขนาดเล็ก และจะเห็นได้ว่าตัวอย่างขนาดเล็ก และขนาดใหญ่ ให้ค่าประมาณ p – value ไม่แตกต่างกัน ในแต่ละค่าพารามิเตอร์

สรุปผลการวิจัย

แม้ว่าตัวอย่างจะมีขนาดเล็ก และค่าพารามิเตอร์ P มีค่าเข้าใกล้ 0 นั้น ซึ่งในการทดลองกลุ่มแบบแบร์นูลลี (Bernoulli Trial) กระทำซ้ำกัน n ครั้งอย่างเป็นอิสระต่อกันด้วยพารามิเตอร์เดียวกันจะนำไปสู่ตัวแปรสุ่มใหม่ที่มีการแจกแจงทวินาม ซึ่งมีลักษณะสมมาตรที่ $P = 0.5$ คือเป็นเหตุการณ์ที่เกิดขึ้นบ่อยที่สุด ส่งผลทำให้เกิดค่าความน่าจะเป็นของการเกิดเหตุการณ์มีค่ามากที่สุด แต่เมื่อ $P = 0.1, 0.9$ ซึ่งไม่ใช่เฉพาะกรณีที่ $P = 0.5$ เท่านั้น ในทุกระดับนัยสำคัญ 0.01, 0.05 ค่าประมาณ p – value ที่ได้จะมากกว่าระดับนัยสำคัญดังกล่าว ซึ่งจะส่งผลถึงการยอมรับสมมติฐาน เนื่องจากการหาค่า p – value บนพื้นฐานสมมติฐานหลักที่เป็นจริงนั้นหมายความว่า ไม่มีโอกาสที่เกิดความผิดพลาดชนิดที่ 1 และถือว่าไม่เป็น

การทดสอบสมมติฐานที่จะปฏิเสธสมมติฐานหลัก
 H_0 ไว้ก่อน ทั้งที่สมมติฐานหลักนั้นเป็นจริง

ข้อเสนอแนะ

สำหรับงานวิจัยนี้ได้ทำการศึกษาค่า p -value ในการทดสอบสมมติฐาน สำหรับพารามิเตอร์ของการแจกแจงแบร์นูลลี ด้วยวิธีบูทสเตรป ซึ่งทำการสุ่มซ้ำแบบใส่คืนจำนวน 1,000 ชุด จึงน่าสนใจศึกษาในกรณีที่มีจำนวนชุดเพิ่มขึ้น และที่สำคัญคือ การที่สามารถสร้างตัวสถิติทดสอบใหม่ ที่ซับซ้อนยิ่งขึ้นแล้วนำไปประยุกต์ใช้กับวิธีบูทสเตรป

เอกสารอ้างอิง

- ขงยุทธ ไชยพงศ์. 2553. เอกสารประกอบ คำ สอนวิชา
ทฤษฎีการอนุมานทางสถิติ สำหรับประชากร
อนันต์. ภาควิชาสถิติ คณะวิทยาศาสตร์
มหาวิทยาลัยเชียงใหม่.
- Casella G. and Berger R.L. 2002. Statistical
Inference. Duxbury Press.
- Garthwaite P. H., Jolliffe I. T. and Jones B.
2002. Statistical Inference. New York:
Oxford University Press.
- Hutson A.D. 2006. Modifying the Exact Test
for a Binomial Proportion and
Comparisons with Other Approaches.
Journal of Applied Statistics. 3: 679-690.
- Krishnamoorthy K. and Peng J. 2007. Some
Properties of the Exact and Score
Methods for Binomial Proportion and
Sample Size Calculation. Communications in
Statistics -Simulation and Computation.
36: 1171-118.
- Lloyd C. J. 2010. How Close are Alternative
Bootstrap P-values?. Statistics &
Probability Letters. 80: 1972-1976.

Lloyd C. J. 2008. Estimated P-values in Discrete
Models: Asymptotic and Non -
asymptotic Effects. MBS Working
Paper. Retrieved January 9, 2011,
from http://works.bepress.com/chris_lloyd/ 15