บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการศึกษาวิจัยทางชีววิทยาและด้านการแพทย์มีหลายกรณีที่พบว่าประชากรที่ศึกษามีลักษณะหายาก (Rare Case) หรือพบได้น้อยมากดังเช่น สตรีที่ตั้งครรภ์เป็นแฝดสาม สตรีที่ตั้งครรภ์แล้วเป็นโรคหัวใจ ผู้ป่วยโรคผิวหนังแข็ง และผู้ป่วยโรคเอฟโอพี (โรคกล้ามเนื้อหรือเอ็นกลายเป็นกระดูก)

โดยแท้จริงแล้วประชากรเช่นนี้เป็นมนุษย์ที่เกิดโรคหรือความผิดปกติที่เกิดขึ้นด้วยความน่าจะเป็นที่น้อยมากหรือเข้าใกล้ศูนย์ นั่นคือการเกิดโรคที่สนใจนั้นพบได้ยากในกลุ่มประชากร หากต้องการประมาณค่าความน่าจะเป็นดังกล่าว ภายใต้มุมมองของประชากรอนันต์ก็ย่อมหมายถึงการประมาณค่าพารามิเตอร์ p ของการแจกแจงแบบแบร์นูลลีนั่นเอง

ในการประยุกต์ทฤษฎีการอนุมานทางสถิติโดยปกติการประมาณค่าพารามิเตอร์ที่ต้องการหากดำเนินการเก็บรวบรวมข้อมูลโดยทำการกำหนดขนาดตัวอย่าง n แล้วทำการทดลองหรือเฝ้าสังเกตจนได้ข้อมูลตัวอย่าง $X_1,...,X_n$ ย่อมนำไปสู่การประมาณค่าพารามิเตอร์ที่ผิดพลาดเพราะค่าข้อมูลตัวอย่าง $X_1,...,X_n$ ย่อมเป็น 0 เกือบทั้งหมดอันเนื่องมาจากคุณสมบัติของประชากรที่หายากนั่นเอง

เพื่อแก้ไขปัญหาที่เกิดขึ้น จึงได้มีการพัฒนาวิธีการเก็บข้อมูลในลักษณะของการสุ่มตัวอย่างแบบผกผัน(Inverse Sampling) โดยแทนที่จะกำหนดขนาดตัวอย่างให้เป็นค่าคงที่นั้นกลับเป็นการกำหนดให้ทำการเก็บรวบรวมข้อมูลจนได้รับหน่วยตัวอย่างที่มีคุณลักษณะตามต้องการจนครบตามจำนวนที่กำหนด แล้วจึงนำข้อมูลที่ได้รับดังกล่าวไปทำการประมาณค่าพารามิเตอร์ที่ต้องการทั้งในลักษณะแบบจุดและแบบช่วง

ภายใต้การสุ่มตัวอย่างแบบผกผันนี้ตัวประมาณแบบภาวะน่าจะเป็นสูงสุดสำหรับทำการประมาณค่าพารามิเตอร์ p มีคุณสมบัติที่อิงฟังก์ชันความน่าจะเป็นแบบทวินามลบ (Negative Binomial) ที่ต่างไปจากกรณีของการสุ่มตัวอย่างปกติโดยตรง (Direct Sampling) ที่ตัวประมาณแบบภาวะน่าจะเป็นสูงสุดมีคุณสมบัติที่อิงฟังก์ชันความน่าจะเป็นแบบทวินาม (Binomial) ดังนั้นตัวประมาณค่าพารามิเตอร์แบบช่วงย่อมถูกสร้างขึ้นด้วยฟังก์ชันความน่าจะเป็นที่แตกต่างกัน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

โดยพื้นฐานการประมาณค่าแบบช่วงนั้นมีการสร้างและพัฒนามาจากการประมาณค่าแบบจุดที่มีโอกาสได้ค่าประมาณเท่ากับค่าพารามิเตอร์ที่แท้จริงนั้นมีค่าเป็นศูนย์ซึ่งเป็นข้อด้อยของการประมาณค่าแบบจุด จึงเป็นเหตุผลอีกประการที่นำไปสู่การสร้างตัวประมาณค่าแบบช่วงขึ้นโดยการประมาณค่าแบบช่วง จะประกอบด้วยตัวประมาณแบบจุด 2 ตัว นั่นคือ ขีดจำกัดบน $U(X_1, X_2, ..., X_n)$ และขีดจำกัดล่าง $L(X_1, X_2, ..., X_n)$ ภายใต้ระดับความเชื่อมั่น 100 $(1-\alpha)$% ที่กำหนด

สำหรับการสร้างตัวประมาณแบบช่วงแรกเริ่มมาจากการสร้างช่วงความเชื่อมั่นด้วยวิธีความน่าจะเป็นสะสม (Cumulative Probability Method) ของฟังก์ชันของตัวอย่างสุ่ม ซึ่งฟังก์ชันที่เหมาะสมที่สุดควรมีคุณสมบัติเป็นสถิติพอเพียง (Sufficient Statistics) หรือการสร้างโดยใช้ส่วนกลับของสถิติทดสอบ (Inverse of Test Statistics) หรือการใช้ปริมาณหมุน (Pivotal Quantity)

การประมาณค่าแบบช่วงสำหรับค่าพารามิเตอร์ p จากการเก็บรวบรวมข้อมูลตัวอย่างแบบผกผันที่อิงฟังก์ชันความน่าจะเป็นแบบทวินามลบนั้น ได้รับการศึกษามาเป็นเวลานานและแพร่หลาย ดังเช่น Casella และ Berger ในปี ค.ศ. 2002 ได้ศึกษาถึงความสัมพันธ์ระหว่างการแจกแจงทวินามและทวินามลบทำให้ได้ช่วงความเชื่อมั่นที่แท้จริง (Exact Confidence Interval: EX) สำหรับสัดส่วนทวินามลบ และ Lui ในปี ค.ศ. 2004 ได้พิจารณาถึงช่วงความเชื่อมั่นของวาล์วสำหรับสัดส่วนทวินามลบซึ่งขึ้นอาศัยวิธีการภาวะน่าจะเป็นสูงสุด (Minimum Likelihood Estimator : MLE) รวมทั้งตัวประมาณที่ไม่เอนเอียงของ p อย่างช่วงความเชื่อมั่นของวาล์วโดยอาศัยวิธีความแปรปรวนต่ำสุดอย่างเอกรูป (Wald-type confidence interval based on UMVUE: WU) และต่อมาได้มีการสร้างช่วงความเชื่อมั่นอื่นๆ ขึ้นมาใหม่ Tian, Tang Ng และ Chan ในปี ค.ศ. 2009 สร้างช่วงความเชื่อมั่นโดยวิธีสกอร์ (Score confidence interval: SC) วิธีอาศัยอัตราส่วนภาวะน่าจะเป็น (Likelihood ratio based on confidence interval: LR) และวิธีอาศัยการประมาณแซดเดิลพอยด์ (Saddlepoint approximation based confidence interval: SP)

ผู้ศึกษาเห็นว่าการสร้างตัวประมาณแบบช่วงภายใต้แนวทางที่อิงการคำนวณด้วยคอมพิวเตอร์เป็นวิธีที่สะดวก และง่ายขึ้นซึ่งก็คือ วิธีบูทแสตรป (Bootstrap Method) เป็นวิธีการที่ใช้กันอย่างแพร่หลายสำหรับสร้างช่วงความเชื่อมั่นซึ่งหลักการสุ่มของวิธีบูทแสตรปคือ การสุ่มตัวอย่างขนาด r แบบใส่คืนจากตัวอย่างชุดเดิมซ้ำๆ กันจำนวน B ครั้ง โดยมีแนวคิดว่าหากการแจกแจงตัวอย่างสุ่มเกิดจากค่าสถิติของตัวอย่างหลายๆ ชุด ดังนั้นวิธีบูทแสตรป ที่เกิดจากการสุ่มหลายๆ ครั้งจึงเป็นการจำลองการแจกแจงของตัวอย่างสุ่มของประชากรจากตัวอย่างเพียงชุดเดียว ในการประมาณค่าความแปรปรวนของค่าสถิติ จะทำการคำนวณจากวิธีบูทแสตรปเกิดจากค่าสถิติโดยตรง แทนที่จะทำการคำนวณโดยตรงจากสูตรทางคณิตศาสตร์ซึ่งยุ่งยาก

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ในการศึกษาครั้งนี้ผู้ศึกษาจึงได้ทำการศึกษาการประมาณค่าแบบช่วงสำหรับค่าสัดส่วนทวินามลบโดยวิธีบูทแสตรป และทำการศึกษาเปรียบเทียบกับอีก 4 วิธีคือ วิธีแบบวาล์วโดยอาศัยวิธีภาวะน่าจะเป็นสูงสุด วิธีแบบวาล์วโดยอาศัยวิธีความแปรปรวนต่ำสุดอย่างเอกรูป วิธีอาศัยการแจกแจงแบบไคสแควร์ และวิธีสกอร์ โดยใช้การจำลองข้อมูล ทั้งนี้ เกณฑ์ที่ใช้ในการตรวจสอบความผิดพลาด และความเชื่อถือได้ของวิธีการประมาณค่าแบบช่วงทั้ง 5 วิธี จะพิจารณาจากค่าความน่าจะเป็นของการครอบคลุมค่าพารามิเตอร์ (Coverage Probability) และค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น(Average Length) เพื่อเป็นแนวทางในการจัดทำข้อสรุปในการเลือกใช้วิธีการประมาณค่าที่ถูกต้อง และเหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาวิธีการประมาณค่าแบบช่วงสำหรับค่าสัดส่วนทวินามลบโดยวิธีบูทแสตรป

2. เพื่อเปรียบเทียบความแตกต่างกันของประสิทธิภาพช่วงความเชื่อมั่นสำหรับค่าสัดส่วนทวินามลบ วิธีแบบวาล์วโดยอาศัยวิธีภาวะน่าจะเป็นสูงสุด วิธีแบบวาล์วโดยอาศัยวิธีความแปรปรวนต่ำสุดอย่างเอกรูป วิธีอาศัยการแจกแจงแบบไคสแควร์ และวิธีสกอร์

3. เพื่อจัดทำข้อสรุปในการเลือกใช้วิธีการประมาณค่าแบบช่วงที่เหมาะสมสำหรับค่าสัดส่วนทวินามลบในแต่ละสถานการณ์

1.3 แผนการวิจัย

1.ซึ่ง Y แทนจำนวนครั้งของการเกิดคุณลักษณะที่ไม่สนใจก่อนที่จะเกิดคุณลักษณะที่สนใจครั้งที่ r มีพารามิเตอร์ r และ p

2. กำหนดจำนวนครั้งของการเกิดคุณลักษณะที่สนใจ (r) เท่ากับ 1, 3, 5, 7, 10, 20, 30 และ 50

3. กำหนดค่าความน่าจะเป็นในการเกิดคุณลักษณะที่สนใจ (p) เท่ากับ 0.001, 0.01, 0.1 และ 0.5

4. กำหนดระดับความเชื่อมั่น (1- α)100% ที่ทำการวิจัยเท่ากับ 95% และ 99%

5. ทำการทดลองซ้ำ 1,000 ครั้งในแต่ละสถานการณ์ที่กำหนด

6. สำหรับการสร้างช่วงความเชื่อมั่นสำหรับพารามิเตอร์ p โดยวิธีบูทแสตรปจะทำการสุ่มตัวอย่างซ้ำแบบใส่คืน จำนวน 1,000 ครั้ง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

1.4 ขั้นตอนการดำเนินการวิจัย

สำหรับขั้นตอนการดำเนินการวิจัยมีดังต่อไปนี้

1. จำลองข้อมูลตามสถานการณ์ต่าง ๆ ที่กำหนดไว้ในแผนการวิจัย

2. คำนวณช่วงความเชื่อมั่นด้วยวิธีการประมาณค่าแบบช่วงทั้ง 5 วิธี ได้แก่

2.1 วิธีแบบวาล์วโดยอาศัยวิธีภาวะน่าจะเป็นสูงสุด

2.2 วิธีแบบวาล์วโดยอาศัยวิธีความแปรปรวนต่ำสุดอย่างเอกรูป

2.3 วิธีอาศัยการแจกแจงแบบไคสแควร์

2.4 วิธีสกอร์

2.5 วิธีบูทแสตรป

3. ดำเนินการตามข้อ 1-2 ซ้ำ จำนวน 1,000 ครั้งในแต่ละสถานการณ์

4. พิจารณาว่าช่วงความเชื่อมั่นที่ได้ในข้อ 2 คลุมค่าพารามิเตอร์ p หรือไม่

5. คำนวณความกว้างของช่วงความเชื่อมั่น

6. คำนวณค่าความน่าจะเป็นของการครอบคลุมค่าพารามิเตอร์ p และค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ได้จากวิธีการประมาณค่าแบบช่วงทั้ง 5 วิธี

7. เปรียบเทียบค่าความน่าจะเป็นของการครอบคลุมค่าพารามิเตอร์ p ที่ได้จากวิธีการประมาณค่าแบบช่วงทั้ง 5 วิธี กับระดับความเชื่อมั่นที่กำหนด

8. เปรียบเทียบค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นเฉพาะวิธีการประมาณค่าแบบช่วงที่ให้ค่าประมาณความน่าจะเป็นของการครอบคลุมค่าพารามิเตอร์ p ไม่น้อยกว่าระดับความเชื่อมั่นที่กำหนด

9. สรุปผลการวิจัย

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ทราบถึงวิธีการประมาณค่าแบบช่วงสำหรับค่าพารามิเตอร์ p จากข้อมูลตัวอย่างแบบผกผันโดยวิธีบูทแสตรป

2. เป็นแนวทางในการเลือกใช้วิธีการประมาณค่าแบบช่วงสำหรับการสำหรับค่าพารามิเตอร์ p จากข้อมูลตัวอย่างแบบผกผันได้อย่างเหมาะสมในแต่ละสถานการณ์

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved