บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การประมาณค่าพารามิเตอร์เป็นวิธีการทางสถิติที่มีความสำคัญมากวิธีหนึ่ง โดยเป็นวิธีที่ใช้ในการประมาณค่าคุณลักษณะต่าง ๆ ของประชากรที่เรียกว่าพารามิเตอร์ โดยอาศัยข้อมูลตัวอย่างในงานวิจัยและการศึกษาต่าง ๆ มีการนำการประมาณค่าพารามิเตอร์มาใช้กันอย่างแพร่หลาย การประมาณค่าพารามิเตอร์แบ่งออกเป็น 2 ประเภท คือ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation) การประมาณค่าแบบจุด คือ การใช้ค่าสถิติจากข้อมูลตัวอย่างเพียงค่าเดียวในการประมาณค่าพารามิเตอร์ เช่น การใช้ค่าเฉลี่ยตัวอย่างในการประมาณค่าเฉลี่ยประชากร เป็นต้น ส่วนการประมาณค่าแบบช่วงคือ การประมาณค่าพารามิเตอร์ด้วยตัวเลขที่เป็นช่วง โดยคาดว่าช่วงที่ประมาณได้นี้จะครอบคลุมค่าพารามิเตอร์ สำหรับการประมาณค่าแบบจุดเป็นการประมาณค่าโดยใช้ค่าประมาณเพียงค่าเดียว ดังนั้นหากค่าประมาณจากตัวอย่างสุ่มมีความแตกต่างจากประชากรมาก การประมาณค่าแบบจุดจะมีประสิทธิภาพต่ำและเสี่ยงต่อความผิดพลาดสูงเนื่องจากไม่มีการกำหนดขนาดของความผิดพลาดไว้ ในขณะที่การประมาณแบบช่วงจะครอบคลุมค่าพารามิเตอร์ที่ต้องการประมาณได้มากกว่าการประมาณค่าแบบจุด โดยอาศัยการแจกแจงความน่าจะเป็นของตัวประมาณค่าที่ได้จากตัวอย่างสุ่มในการสร้างช่วงของการประมาณซึ่งทำให้ทราบว่าในการประมาณแต่ละครั้งมีความถูกต้องร้อยละเท่าใด

การวัดการกระจาย (Measure of Dispersion) เป็นวิธีการทางสถิติที่ใช้อธิบายลักษณะการกระจายของข้อมูล หากข้อมูลชุดหนึ่งประกอบด้วยค่าสังเกตที่มีค่าแตกต่างกันมากแสดงว่าข้อมูลชุดนั้นมีการกระจายมาก หากข้อมูลชุดนั้นมีค่าสังเกตที่มีค่าแตกต่างกันน้อยแสดงว่าข้อมูลชุดนั้นมีการกระจายน้อย ค่าสถิติที่ใช้สำหรับวัดการกระจายของข้อมูลมีอยู่ด้วยกันหลายชนิด เช่น พิสัย (Range) ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) และสัมประสิทธิ์การแปรผัน (Coefficient of Variation หรือ CV) เป็นต้น อย่างไรก็ตาม พิสัยและค่าเบี่ยงเบนมาตรฐาน เป็นค่าที่ใช้วัดการกระจายที่มีหน่วยเดียวกับค่าของข้อมูล ทำให้ไม่สามารถใช้ค่าเหล่านี้ในการเปรียบเทียบการกระจายของข้อมูลตั้งแต่สองชุดที่มีหน่วยต่างกันได้ ในขณะที่สัมประสิทธิ์การแปรผันเป็นค่าที่ใช้วัดการกระจายที่ไม่มีหน่วย จึงสามารถใช้ในการเปรียบเทียบการกระจายของข้อมูลตั้งแต่สองชุดขึ้นไปที่มีหน่วยแตกต่างกันได้ โดยข้อมูลชุดที่มีสัมประสิทธิ์การแปรผันสูงมีการกระจายมากกว่า

ข้อมูลชุดที่มีสัมประสิทธิ์การแปรผันต่ำ ด้วยคุณสมบัตินี้จึงทำให้สัมประสิทธิ์การแปรผันเป็นที่นิยมใช้ในการอธิบายการกระจายของข้อมูลในหลาย ๆ สาขาวิชา เช่น ปริมาณน้ำมันที่ใช้ในเครื่องยนต์ในสาขาวิศวกรรมศาสตร์ ปริมาณเซลล์เม็ดเลือดขาวที่อยู่ในเลือด ในสาขาวิทยาศาสตร์ เป็นต้น อย่างไรก็ตามจากการศึกษาของ Bonett (2006a) พบว่าในกรณีที่ข้อมูลของประชากรไม่มีการแจกแจงปรกติ สัมประสิทธิ์การแปรผันไม่ใช่ค่าวัดการกระจายที่เหมาะสมเนื่องจากในการคำนวณค่าสัมประสิทธิ์การแปรผันต้องใช้ค่าเบี่ยงเบนมาตรฐานและค่าเฉลี่ย ซึ่งพารามิเตอร์ดังกล่าวจะเป็นค่าวัดที่เหมาะสมเมื่อข้อมูลของประชากรมีการแจกแจงปรกติ โดยค่าที่ใช้วัดการกระจายที่เหมาะสมกว่าสัมประสิทธิ์การแปรผันในกรณีนี้ คือสัมประสิทธิ์การแปรผันควอร์ไทล์ (Coefficient of Quartile Variation หรือ CQV)

สัมประสิทธิ์การแปรผันควอร์ไทล์ คือค่าที่ใช้วัดการกระจายของชุดข้อมูล ที่คำนวณได้จาก

$$CQV = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Q_1 คือ ควอร์ไทล์ที่ 1 หรือ เปอร์เซ็นต์ไทล์ที่ 25 ของข้อมูลประชากร

Q_3 คือ ควอร์ไทล์ที่ 3 หรือ เปอร์เซ็นต์ไทล์ที่ 75 ของข้อมูลประชากร

Bonett (2006a) ได้สร้างช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์และพบว่าช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์มีประสิทธิภาพดี เมื่อข้อมูลประชากรไม่มีการแจกแจงปรกติ

Efron and Tibshirani (1994) ได้เสนอวิธีการประมาณค่าพารามิเตอร์โดยใช้เทคนิคการสุ่มซ้ำ (Resampling Technique) จากตัวอย่างสุ่มที่มีเพียงชุดเดียวโดยใช้การชักตัวอย่างแบบคืนที่ (Sampling with Replacement) ทำให้ได้ชุดตัวอย่างที่เป็นอิสระกันขนาด n จำนวน เมื่อ n คือขนาดตัวอย่าง B ชุด เมื่อ B คือจำนวนชุดตัวอย่าง และเรียกชุดตัวอย่างเหล่านี้ว่าตัวอย่างบูตสแตรป (Bootstrap Sample) จากนั้นจึงประมาณค่าพารามิเตอร์ที่สนใจด้วยตัวอย่างบูตสแตรป ซึ่งในกรณีที่ไม่ทราบลักษณะการแจกแจงของประชากรหรือการแจกแจงของประชากรไม่มีการแจกแจงปรกติ วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีบูตสแตรปมีความเหมาะสม เนื่องจากวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีบูตสแตรป เป็นวิธีการหาตัวประมาณค่าด้วยวิธีสถิติศาสตร์ไม่อิงพารามิเตอร์ (Nonparamatrics Statistics) ที่ไม่มีข้อจำกัดเกี่ยวกับลักษณะการแจกแจงของประชากร

นอกจากนี้ยังได้นำเสนอวิธีการประมาณช่วงความเชื่อมั่นสำหรับพารามิเตอร์ต่าง ๆ ด้วยเทคนิคการสุ่มซ้ำ เช่น วิธีการประมาณค่าคลาดเคลื่อนมาตรฐานของตัวประมาณค่าต่าง ๆ ที่ได้จากการสุ่มซ้ำ และนำมาสร้างช่วงความเชื่อมั่นสำหรับพารามิเตอร์ที่ต้องการประมาณ โดยเรียกว่า

การประมาณค่าคลาดเคลื่อนมาตรฐานด้วยวิธีบูตสแตรป (Bootstrap Estimate for Standard Error) และอีกวิธีที่นิยม คือวิธีบูตสแตรปที (Bootstrap-t) ซึ่งใช้ในการประมาณช่วงความเชื่อมั่นสำหรับพารามิเตอร์โดยการประมาณค่าคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเช่นเดียวกับวิธีข้างต้น แตกต่างกันที่วิธีข้างต้นใช้ z จากการแจกแจงปรกติมาตรฐาน แต่วิธีบูตสแตรปทีจะทำการประมาณการแจกแจงของตัวประมาณค่า (t^*) จากตัวอย่างบูตสแตรป ที่ได้จากการสุ่มซ้ำ

พรรณนา เอี่ยมสุวรรณ์ (2555) ได้เสนอวิธีการประมาณช่วงความเชื่อมั่นสำหรับค่าพิสัยควอร์ไทล์ โดยใช้วิธีการประมาณค่าคลาดเคลื่อนมาตรฐานด้วยวิธีบูตสแตรปและวิธีบูตสแตรปที ในการประมาณช่วงความเชื่อมั่นสำหรับค่าพิสัยควอร์ไทล์ จากการจำลองข้อมูลจากประชากรที่มีการแจกแจงรูปแบบต่าง ๆ ด้วยขนาดตัวอย่าง 5 ถึง 100 พบว่าวิธีบูตสแตรปทีมีประสิทธิภาพที่ดียกเว้นในกรณีประชากรมีการแจกแจงบีต้า เมื่อ α =0.5, β =0.5 การแจกแจงแกมมา เมื่อ α =0.5, β =1 และการแจกแจงล็อกนอร์มอล เมื่อ $\mu = 0$, $\sigma^2 = 1$ ส่วนวิธีประมาณค่าเบี่ยงเบนมาตรฐานของตัวประมาณค่าพิสัยควอร์ไทล์ด้วยวิธีบูตสแตรป นั้นมีประสิทธิภาพที่ดียกเว้นกรณีที่ประชากรมีการแจกแจงบีต้า เมื่อ α =0.5, β =0.5

ในงานวิจัยนี้ผู้วิจัยสนใจพัฒนาวิธีการประมาณช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์โดยใช้วิธีบูตสแตรป และบูตสแตรปที จากแนวคิดที่กล่าวมาข้างต้น

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการประมาณช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์ 4 วิธีคือ

- วิธีบูตสแตรปเปอร์เซ็นต์ไทล์ (Percentile-Bootstrap)
- วิธีบูตสแตรป Z (Bootstrap-Z)
- วิธีบูตสแตรปที (Bootstrap-t)
- วิธีบูตสแตรปทีร่วมกับวิธีของ Bonett (Bonett-Bootstrap)

2. เปรียบเทียบประสิทธิภาพของวิธีการประมาณช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์ทั้ง 4 วิธี และวิธีของ Bonett

มหาวิทยาลัยบูรพา Burapha University

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ทราบถึงวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์ที่เหมาะสม เมื่อข้อมูลของประชากรมีการแจกแจงความน่าจะเป็นแบบปรกติและเบ้ขวา

2. เป็นแนวทางในการศึกษาเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์กับวิธีอื่น ๆ

ขอบเขตการวิจัย

ในการศึกษาครั้งนี้กำหนดสถานการณ์ต่าง ๆ เพื่อประมาณช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์ โดยกำหนดค่าพารามิเตอร์ให้การแจกแจงของประชากรมีการแจกแจงปรกติและเบ้ขวาโดยข้อมูลมีค่ามากกว่าหรือเท่ากับ 0 ดังนี้

1. การแจกแจงความน่าจะเป็นของประชากรที่ศึกษา คือ

- การแจกแจงปรกติ เมื่อ $\mu = 4$ และ $\sigma^2 = 1$ สัมประสิทธิ์ความเบ้ คือ 0
- การแจกแจงล็อกนอร์มอล เมื่อ $\mu = 0$ และ $\sigma^2 = 1$ สัมประสิทธิ์ความเบ้ คือ 6.1848
- การแจงแจงแกมมา เมื่อ (α , β) คือ (0.5, 1) (1.5, 1) (6, 1) สัมประสิทธิ์ความเบ้ คือ 2.8284 1.6329 และ 0.8165 ตามลำดับ
- การแจกแจงบีต้า เมื่อ $\alpha = 2$ และ $\beta = 4$ สัมประสิทธิ์ความเบ้ คือ 0.4677

2. ขนาดตัวอย่าง (n)

- ตัวอย่างขนาดเล็ก คือ 10 และ 25
- ตัวอย่างขนาดใหญ่ คือ 50 และ 100

3. สัมประสิทธิ์ความเชื่อมั่น 0.95

4. จำนวนรอบของการสุ่มโดยวิธีบูตสแตรป (B) เท่ากับ 2,000 ครั้ง

5. จำนวนรอบของการจำลองตัวอย่างสุ่มเท่ากับ 50,000 ครั้ง

6. ช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์ที่ศึกษา 5 วิธี ได้แก่

- วิธีของ Bonett (2006)
- วิธีบูตสแตรปเปอร์เซ็นต์ไทล์
- วิธีบูตสแตรป Z
- วิธีบูตสแตรปที
- วิธีบูตสแตรปทีร่วมกับวิธีของ Bonett

7. มีเกณฑ์ในการวัดประสิทธิภาพของวิธีการประมาณช่วงความเชื่อมั่นสำหรับสัมประสิทธิ์การแปรผันควอร์ไทล์โดยพิจารณาจากค่าความน่าจะเป็นครอบคลุม (Coverage Probability) และค่าความยาวเฉลี่ย (Average Length) ของช่วงความเชื่อมั่น

มหาวิทยาลัยบูรพา
Burapha University