



## บทที่ 1

### บทนำ

#### ความสำคัญและความเป็นมาของปัญหา

ในการนำข้อมูลสถิติซึ่งเป็นผลจากการสำรวจไปใช้งานนั้น ปัญหาที่มักมีผู้สนใจถามนักสถิติที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลสถิติก็คือ ผลการสำรวจมีความถูกต้องและเชื่อถือได้มากน้อยเพียงไร การตอบคำถามดังกล่าวผู้วิเคราะห์ข้อมูลและผู้ใช้ผลการวิเคราะห์ข้อมูลควรทำความเข้าใจกับลักษณะที่มาของความคลาดเคลื่อนของข้อมูลที่ได้จากการสำรวจ ซึ่งอาจแบ่งความคลาดเคลื่อนออกเป็น 2 ประเภทคือ

1 ความคลาดเคลื่อนจากการใช้ตัวอย่าง(Sampling error) ได้แก่ ความแปรปรวนในการประมาณตัวเลขซึ่งเกิดจากการเลือกตัวอย่าง เช่น วิธีการเลือกตัวอย่าง ขนาดตัวอย่างและสูตรที่ใช้ในการประมาณ

2 ความคลาดเคลื่อนที่ไม่ได้เกิดจากการใช้ตัวอย่าง(Nonsampling error)ความคลาดเคลื่อนประเภทนี้เกิดขึ้นในขั้นตอนอื่นๆของการสำรวจตัวอย่างนอกเหนือจากแผนการสุ่มตัวอย่าง ดังนั้นความบกพร่องที่อาจเกิดขึ้นนี้อาจเกิดขึ้นในขั้นตอนการวางแผนและการดำเนินงานสำรวจตัวอย่าง ตั้งแต่ขั้นตอนการเลือกหรือสร้างกรอบตัวอย่าง การสร้างสื่อในการเก็บข้อมูล เช่น แบบสอบถาม ความผิดพลาดในข้อมูลที่เก็บได้รวมทั้งการประมวลผลข้อมูล และสาเหตุหนึ่งที่สำคัญของความคลาดเคลื่อนที่ไม่ได้เกิดจากการเลือกตัวอย่างคือ การไม่ตอบกลับ(Nonresponse)ซึ่งเป็นกรณีที่ไม่สามารถได้คำตอบหรือวัดค่าของหน่วยตัวอย่างได้ อาจมีสาเหตุมาจากการหาหน่วยตัวอย่างไม่พบหรือหน่วยตัวอย่างไม่อยู่เมื่อพนักงานสัมภาษณ์ไปสัมภาษณ์ หรือหน่วยตัวอย่างไม่ยอมให้ข้อมูลหรือตอบคำถามหรือตอบไม่ได้ เช่นคำถามที่ถามต้องการความรู้เฉพาะด้านใดด้านหนึ่งซึ่งหน่วยตัวอย่างไม่มีความรู้ด้านนั้นพอที่จะตอบได้ การไม่ตอบเป็นปัญหาที่เกิดขึ้นเสมอในการเก็บข้อมูลปฐมภูมิ โดยเฉพาะอย่างยิ่งในการศึกษาทางสังคมศาสตร์ที่ต้องอาศัยข้อมูลจากบุคคลจะปรากฏปัญหาเรื่องการไม่ตอบกลับอยู่เสมอและอัตราการไม่ตอบกลับค่อนข้างสูงด้วย สำหรับผลสำคัญของการไม่ตอบก็คือ ทำให้หน่วยตัวอย่างที่วัดได้มีจำนวนน้อยกว่าที่กำหนดไว้ซึ่งย่อมกระทบกระเทือนถึงคุณภาพของตัวประมาณเพราะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงขึ้น นอกจากนี้ผลของการไม่ตอบกลับอีกประการหนึ่งคือ ทำให้เกิดความเอนเอียงในตัวประมาณ ทั้งนี้

เพราะ โดยทั่วไปแล้วลักษณะของหน่วยที่ไม่ตอบกับหน่วยที่ตอบมักแตกต่างกัน ดังนั้นในทางปฏิบัติ นักสถิติอาจลดจำนวนผู้ไม่ตอบกลับด้วยวิธีต่างๆ ดังนี้

2.1 กำหนดขั้นตอนในการเก็บรวบรวมข้อมูลให้เหมาะสม เช่น แจกให้ผู้ที่ถูกเป็นตัวอย่างในการสำรวจให้ทราบล่วงหน้า หรือหาวิธีจูงใจให้ผู้ตอบสัมภาษณ์ทราบถึงความสำคัญเพื่อที่จะได้ให้ความร่วมมือมากยิ่งขึ้น

2.2 ควรเลือกช่วงเวลาที่ใช้ในการสำรวจให้เหมาะสม เช่น จะสอบถามชานาควรเป็นหลังฤดูเก็บเกี่ยว ถ้าจะถามนิสิตนักศึกษาควรถามตอนปิดเทอม ส่วนเวลานั้นควรเป็นตอนเย็นหรือหลังจากเลิกงานแล้วหรือวันเสาร์อาทิตย์

2.3 ในการสัมภาษณ์ครั้งแรก ไม่ได้รับคำตอบผู้ที่ทำการสัมภาษณ์ควรไปติดต่อเพื่อทำการสัมภาษณ์อีกเป็นครั้งที่สองหรือครั้งที่สามเพื่อให้ได้ข้อมูลที่ต้องการ

2.4 ถ้าพยายามติดต่อสัมภาษณ์แล้วยังไม่ได้ข้อมูลที่ต้องการควรเลือกตัวอย่างเพิ่มเติมตัวอย่างที่ไม่ได้รับคำตอบนั้น ๆ ซึ่งเป็นวิธีการที่สะดวกที่สุดแต่อาจเป็นวิธีที่ไม่ดีนัก

อย่างไรก็ตามถึงแม้ นักสถิติจะใช้วิธีการวางแผนที่ดีและปรับปรุงระเบียบวิธีการต่างๆ ตลอดจนใช้ความสามารถและพยายามต่างๆ แล้วแต่ก็ยังไม่สามารถทำให้จำนวนผู้ไม่ตอบสัมภาษณ์หมดไปได้ ซึ่งเป็นปัญหาที่นักสถิติหรือผู้ที่ทำการสำรวจจะต้องแก้ไขในขั้นประมวลผลต่อไป และในขั้นประมวลผลนี้ นักสถิติจะต้องหาวิธีการประมาณค่าของสิ่งที่ต้องการศึกษาจากประชากร ซึ่งก็มีวิธีการที่ใช้ความน่าจะเป็นมาช่วยลดความเอนเอียงของตัวประมาณ โดยไม่ต้องให้หน่วยที่ไม่ตอบกลับตอบคำถามอีก นอกจากนี้ปัญหาการไม่ตอบกลับสามารถเกิดขึ้นได้อีกในกลุ่มของตัวอย่างที่ตอบกลับมากคือปัญหาการไม่ตอบกลับบางข้อถามที่สนใจ วิทยานิพนธ์เล่มนี้เสนอการประมาณค่าเฉลี่ยของประชากรเมื่อสุ่มตัวอย่างจากประชากรและตัวอย่างสุ่มตอบแบบสอบถามแต่มีบางหน่วยตัวอย่างไม่ตอบข้อถามที่สนใจ ทั้งนี้โดยทั่วไปแล้วในการวิจัยเรื่องหนึ่งมักจะมีการเก็บรวบรวมข้อมูลของตัวแปรต่าง ๆ หลายตัวแปรไปพร้อมกันเพื่อตอบสนองต่อวัตถุประสงค์ของการวิจัยที่กำหนดไว้และจะพบอยู่เสมอว่าในบรรดาตัวแปรต่างๆ ที่เก็บมานั้นมีตัวแปรบางตัวมีความสัมพันธ์กัน ดังนั้นจึงสามารถใช้ความสัมพันธ์ของตัวแปรเหล่านี้ประมาณค่าพารามิเตอร์ของประชากรซึ่งเป็นการเพิ่มประสิทธิภาพของการประมาณค่าดังกล่าว วิทยานิพนธ์นี้ได้เสนอตัวประมาณค่าเฉพาะกรณีเลือกตัวอย่างสุ่มอย่างง่ายชนิดไม่ใส่คืน



## วัตถุประสงค์ของการวิจัย

1. เสนอวิธีการประมาณค่าเฉลี่ยของประชากร โดยดัดแปลงจากวิธีการของ Isaac Olayiwola Oshungade เมื่อเกิดการไม่ตอบกลับในการสุ่มตัวอย่างอย่างง่าย

2. เปรียบเทียบค่าประมาณค่าเฉลี่ยของประชากรของวิธีซึ่งได้จากการดัดแปลงกับวิธีที่ใช้อยู่เดิมอีก 3 วิธี คือ

- ค่าเฉลี่ยจากการตอบกลับ (Response mean)
- วิธีของแฮนเซน - เฮอวิทซ์ (Hansen-Hurwitz)
- วิธีของฟิลลิป เอส คอทท์ (Phillip s. Kott)

## สมมติฐานของการวิจัย

ค่าเฉลี่ยที่ได้จากวิธีของฟิลลิป เอส คอทท์ ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีอื่น ๆ

## ขอบเขตของการวิจัย

1. พิจารณาค่าประมาณของพารามิเตอร์  $\mu$  เท่านั้น
2. ตัวแปรที่ศึกษาเป็นแบบต่อเนื่อง คือตัวแปรตาม (y) และตัวแปรอิสระ (x)
3. ในการวิจัยครั้งนี้จะจำลองการทดลองในเครื่องคอมพิวเตอร์โดยสร้างประชากรขนาด 2000 ซึ่งมีการแจกแจงแบบปกติ กำหนดสัมประสิทธิ์ความแปรผันของประชากร (c.v.) เท่ากับ 5% 10% 15% 20% และ 30% กล่าวคือกำหนดให้  $(\mu, \sigma, I=1,2,3,4,5)$  เท่ากับ  $(150, 7.5)$ ,  $(150, 15)$ ,  $(150, 22.5)$ ,  $(150, 30)$  และ  $(150, 45)$  ตามลำดับ
4. จากประชากรที่จำลองขึ้นขนาดดังกล่าว สุ่มตัวอย่างร้อยละ 5, 10, 15 และ 20
5. กำหนดอัตราการไม่ตอบกลับ (nonresponse) ร้อยละ 5, 10, 20 และ 40 เนื่องจากอัตราการไม่ตอบกลับเท่ากับ 40 เป็นอัตราการไม่ตอบกลับที่อยู่ในระดับสูงมีค่าสูงกว่านี้มางานวิจัยที่ได้มาอาจขาดความน่าเชื่อถือได้ ดังนั้นจึงพิจารณาอัตราการไม่ตอบกลับในระดับสูงที่อัตราการไม่ตอบกลับเท่ากับ 40
6. ใช้ข้อมูลจากโครงการการทำงานของสตรีกับความต้องการผู้ดูแลเด็กวัยก่อนเข้าโรงเรียนอนุบาลในเขตกรุงเทพมหานคร ระหว่างธันวาคม 2532-มกราคม 2533 งานวิจัยนี้เลือกตัวอย่าง

เฉพาะในเขตกรุงเทพมหานคร เลือกสตรีเป็นตัวอย่างของโครงการ โดยสัมภาษณ์สตรีตัวอย่างได้ 1154 ราย ซึ่งเป็นโครงการที่ดำเนินการโดยสถาบันประชากรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย งานวิจัยนี้คัดลอกข้อมูลบางส่วน สำหรับตัวแปรที่นำมาศึกษาคือ

#### 6.1 การทำงานของสตรี

- การศึกษาสูงสุด
- รายได้ต่อเดือน

#### 6.2 รายได้เฉลี่ยต่อเดือนของสามี

7. การกำหนดเกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพของตัวประมาณค่าพารามิเตอร์ โดยพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าเฉลี่ย ( $MSE(\bar{y}_i)$ ) กล่าวคือ เมื่อคำนวณค่าเฉลี่ย ( $\bar{y}_i$ ) ในแต่ละชุดตัวอย่างแล้ว ( $i$  หมายถึงวิธีการที่  $i$ ) จึงคำนวณหาค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ  $\bar{y}_i$  อีกครั้งหนึ่ง ซึ่งถ้า  $MSE(\bar{y}_i)$  มีค่าน้อยแสดงว่าสามารถประมาณค่าพารามิเตอร์ได้ดี

#### ประโยชน์ของการวิจัย

1. ทำให้ทราบวิธีการที่เหมาะสมและเป็นการเพิ่มประสิทธิภาพของการประมาณค่าพารามิเตอร์ของประชากรในกรณีสุ่มตัวอย่างอย่างง่ายและเกิดปัญหาการไม่ตอบกลับ
2. สามารถนำวิธีการประมาณค่าจากข้อ 1 ไปประยุกต์ใช้กับข้อมูลจริงได้
3. สามารถนำวิธีการประมาณค่าเฉลี่ยประชากรที่ได้ไปประมาณค่ารวมของประชากรและประมาณสัดส่วนของสิ่งที่สนใจของประชากรได้

#### ความหมายของคำต่างๆที่ใช้ในงานวิจัย

การไม่ตอบ หมายถึง การส่งแบบสอบถามทางไปรษณีย์และเกิดปัญหาการไม่ตอบคำถามที่ผู้วิจัยสนใจ