



## บทที่ 2

### ทฤษฎีและระเบียบวิธีที่ใช้ในการวัดความคลาดเคลื่อนของข้อมูล

ในสมัยก่อนจุดมุ่งหมายของการทำสำมะโนก็เพื่อที่จะผลิตข้อมูลที่มีคุณภาพที่เป็นไปได้และต่ำสุดเท่านั้น โดยไม่ได้คำนึงถึงความถูกต้องแม่นยำของข้อมูลซึ่งจะมีผลต่อผู้นำข้อมูลไปใช้ แต่ในปัจจุบันได้ให้ความสนใจในการวัดความคลาดเคลื่อนของข้อมูลที่ได้จากการทำสำมะโน แต่อย่างไรก็ตามเป็นที่ยอมรับกันว่าการที่จะได้ข้อมูลจากสำมะโนครบบริบูรณ์นั้นเป็นสิ่งที่ เป็นไปไม่ได้ ("Perfect Census is Impossible") อย่างไรก็ตามข้อมูลที่ได้นี้ก็ยังมีประโยชน์อยู่ ถ้าสามารถทราบขนาดของความคลาดเคลื่อน ดังนั้นวิธีการวัดความคลาดเคลื่อนของข้อมูลจากสำมะโนจึงเป็นสิ่งที่จำเป็นและเป็นวิธีเดียวเท่านั้นที่จะช่วยแสดงถึงความถูกต้องแม่นยำของข้อมูลจากสำมะโน

#### วัตถุประสงค์ของการวัดความคลาดเคลื่อนของสำมะโน

1. วัดความถูกต้องแม่นยำ (Measure accuracy) เพื่อดูว่าข้อมูลที่ได้จากการทำสำมะโนถูกต้องแค่ไหน ซึ่งความรู้เกี่ยวกับความถูกต้องแม่นยำของข้อมูลนี้จะ เป็นแนวทางให้ผู้ที่จะใช้ข้อมูลสามารถนำไปใช้ได้ถูกต้อง
2. ระบุแหล่งที่มาของความคลาดเคลื่อน (Identify sources of error) ทำให้สามารถทราบว่าความคลาดเคลื่อนของข้อมูลที่ได้นั้น เกิดมาจากแหล่งใดบ้าง ซึ่งจะเป็นประโยชน์ต่อผู้ผลิตข้อมูลในการที่จะพยายามลดความคลาดเคลื่อนตามแหล่งที่เกิดนั้น ๆ และก็จะ เป็นประโยชน์ต่อผู้ใช้สถิติช่วยในการที่จะใช้ข้อมูลอย่าง ระมัดระวังว่าที่ใดที่มีความคลาดเคลื่อนบ้าง
3. เพื่อเป็นแนวทางในการที่จะทำสำมะโนในครั้งต่อ ๆ ไป เป็นการช่วยให้ทั้งผู้ผลิตและผู้ได้ข้อมูลสถิติพิจารณาในการวางแผนการทำสำมะโนหรือสำรวจในครั้งต่อไปเพื่อที่จะได้ข้อมูลที่ดียิ่งกว่าเดิม

### ชนิดของความคลาดเคลื่อน

ความคลาดเคลื่อนของข้อมูลที่เกิดจากการเก็บรวบรวมข้อมูล มีอยู่ 2 ประเภท คือ

1. ความคลาดเคลื่อนที่เกิดจากการใช้ตัวอย่าง (Sampling Error) จะเกิดเฉพาะในการทำสำรวจเท่านั้น

2. ความคลาดเคลื่อนที่ไม่ได้เกิดเพราะการใช้ระเบียบวิธีตัวอย่าง (Non-Sampling Error) จะเกิดทั้งในการทำสำมะโนและการทำสำรวจ

ในวิทยานิพนธ์ฉบับนี้จะศึกษาเฉพาะความคลาดเคลื่อนที่ไม่ได้เกิดเพราะการใช้ระเบียบวิธีตัวอย่าง ตามปกติความถูกต้องของข้อมูลสำมะโนขึ้นอยู่กับ

1. คุณภาพของงานสนาม
2. การควบคุมคุณภาพของงานประมวลผล
3. การประมาณค่าเมื่อไม่ได้รับคำตอบ (Imputation for nonresponse)

สาเหตุประการแรกเป็นสิ่งสำคัญยิ่ง เป็นเรื่องที่เราจะต้องวัดผล

ความคลาดเคลื่อนที่ไม่ได้เกิด เพราะการใช้ระเบียบวิธีตัวอย่าง มี 2 อย่าง คือ

1. ความคลาดเคลื่อนของคัมภีร์รวม เป็นความคลาดเคลื่อนจากการนับจดจำนวนรวมของประชากรและเคหะสถาน ซึ่งอาจจะเกิดจากความบกพร่องในการทำแผนที่เขตสำมะโน หรือเกิดจากความบกพร่องของงานสนาม ได้แก่ ความบกพร่องของพนักงานแจงนับ (Enumerator) และผู้คุมงานสนาม (Supervisor) หรืออาจจะเนื่องจากผู้ตอบไม่เข้าใจคำถามดีพอ ซึ่งลักษณะของความคลาดเคลื่อนของคัมภีร์มีดังนี้ คือ

ก. การตกนับจด (Omissions) เป็นการนับจดหน่วยแจงนับ (Enumeration Unit) ใดที่ผิดคือครัวเรือนและประชากรมาไม่ครบ ทำให้ได้จำนวนข้อมูลไม่ครบ

ข. การนับจดซ้ำ (Duplications) เป็นการนับจดหน่วยแจงนับเดียวกันมามากกว่าหนึ่งครั้ง ทำให้ได้จำนวนข้อมูลเกินจากความเป็นจริง

ค. การนับจดเกิน (Erroneous inclusions) เป็นการนับจดหน่วยแจงนับหรือลักษณะบางอย่างที่ไม่ควรนับรวมเข้ามา ซึ่งการนับจดเกินนี้อาจจะเกิดจาก

1. ความไม่ชัดเจนของ เส้นแบ่ง เขตสามะโน ทำให้ระบุขอบเขตท้องที่ผิด
2. เด็กที่เกิดหลัง วันสามะโน
3. คนที่ตายก่อนวันสามะโน
4. คนที่ไม่เข้าข่าย เช่น เพิ่งมาอยู่ ที่ วันสามะโน

2. ความคลาดเคลื่อนของการแจ้งลักษณะ เป็นความคลาดเคลื่อนจากการแจ้งนับลักษณะ บางอย่างของประชากรหรือ คณะสถานส่วนมากมักจะเป็นความคลาดเคลื่อนที่เกิดจากคำตอบที่เกี่ยวข้องกับลักษณะบางอย่าง เช่น อายุ เพศ สถานภาพสมรส อาชีพ ฯลฯ ซึ่งอาจแยกความคลาดเคลื่อนชนิดนี้ได้เป็น 2 แบบ

ก. ตอบผิด (Response errors) ซึ่งเกิดได้จากสาเหตุต่าง ๆ

1. ความไม่เข้าใจของผู้ถามหรือผู้ตอบ
2. ผู้ตอบคำถามไม่เต็มใจที่จะตอบคำถามตามความเป็นจริง
3. ผู้ตอบจำผิด
4. พนักงานบันทึกผิดโดยตั้งใจหรือไม่ตั้งใจก็ตาม

ข. ไม่ได้คำตอบ (Non-Response): อาจเป็นเพราะผู้ตอบจำไม่ได้ หรือพนักงานเดินเลื้อ ไม่ได้ถาม หรือไม่ได้บันทึกมา

เทคนิคต่าง ๆ ที่ใช้วัดความคลาดเคลื่อนของสามะโน

เทคนิคต่าง ๆ ที่ใช้วัดความคลาดเคลื่อนของสามะโนมีดังนี้ คือ

1. การวิเคราะห์ความแนบเนียนภายใน (Analysis of internal consistency check)
2. การวิเคราะห์ทางด้านประชากรศาสตร์หรือการเปรียบเทียบข้อมูลที่ได้กับแหล่งอื่น (Demographic analysis or Comparison with external figure)
3. การประเมินผลความคลาดเคลื่อนโดยวิธีตรง หรือการประมาณค่าจากสองระบบ (Direct methods of evaluation or Dual system estimation)

#### 4. วิธีทางสถิติ (Statistical Methods)

##### 1. การวิเคราะห์ความแนบเนียนภายใน (Analysis of internal Consistency Check)

เป็นวิธีการที่จะนำมาใช้หลังจากการทำสำมะโนเสร็จเรียบร้อยแล้ว และข้อมูลที่ได้ก็ถูกพิมพ์ออกเป็นตารางแล้ว เพราะวิธีการนี้ไม่ต้องออกงานสำเนา จุดมุ่งหมายของวิธีการนี้เพื่อตรวจสอบคุณภาพของข้อมูลจากสำมะโนว่ามีความแนบเนียนกับสิ่งที่กำหนดให้หรือความสัมพันธ์บางอย่างหรือไม่ ตัวอย่าง เช่น

- ตรวจสอบความสัมพันธ์เหตุผลสมผลของความสัมพันธ์ระหว่างรายการต่าง ๆ ในสำมะโน เช่น อาชีพบางชนิดจะมีความสัมพันธ์กับประเภทของอุตสาหกรรม อายุกับระดับการศึกษา เป็นต้น ตัวอย่าง เช่น จำนวนคนในอุตสาหกรรมที่ถูกจำแนกอยู่ในประเภทเกษตร ไม่ควรที่จะต่างกันมากกับจำนวนคนที่มีอาชีพเกษตรกร และก็ไม่ควรที่จะมีคนอายุ 17 ปี ที่มีตำแหน่ง เป็นถึงผู้อำนวยการของธนาคาร หรือเป็นผู้ที่จบมหาวิทยาลัย

- ตรวจสอบตารางต่าง ๆ ว่ามีการกระจายที่สมเหตุสมผลหรือไม่ เช่น ดูจากจำนวนเพศหญิงที่แต่งงานในขณะอายุน้อยจะมีจำนวนมากกว่า เพศชาย หรือก็คือเพศชายที่มีอายุน้อยจะเป็นโสดมากกว่า เพศหญิง และอาจจะตรวจสอบได้จากจำนวนคนที่หย่าหรือเป็นหม้ายในกลุ่มผู้อายุน้อยจะมีจำนวนน้อยกว่ากลุ่มอื่น

##### ข้อดีและข้อเสียของการใช้การวิเคราะห์ความแนบเนียนภายใน

ข้อดีของการใช้การวิเคราะห์ความแนบเนียนภายใน คือ

1. ไม่ต้องออกไปทำงานสำเนา
2. เสียค่าใช้จ่ายน้อยกว่าการทำสำเนาหลังจากการแจกนับ

ข้อเสียของการใช้การวิเคราะห์ความแนบเนียนภายใน

1. การใช้วิธีนี้จะจำกัดเฉพาะในบางกลุ่มของคนและหรือบางรายการของข้อมูลเท่านั้น ไม่สามารถที่จะขยายขอบเขตของการประเมินค่าความคลาดเคลื่อนออกไปตามที่ต้องการได้

2. วิธีการนี้วัดได้เฉพาะความคลาดเคลื่อนรวมเท่านั้น ไม่สามารถแยกวัดออกเป็นความคลาดเคลื่อนของคัมรวม หรือความคลาดเคลื่อนของการแจ้งลักษณะ

3. วิธีการนี้จะบอกแต่ว่า มีความคลาดเคลื่อนเกิดขึ้นเท่านั้น แต่ไม่ได้บอกว่ามี ความคลาดเคลื่อนเกิดขึ้นมากน้อยแค่ไหน

2. การวิเคราะห์ทางด้านประชากรค่าสิทธิ์หรือการเปรียบเทียบข้อมูลที่ได้กับแหล่งอื่น (Demographic analysis หรือ Comparison with external figure)

วิธีการนี้จะเป็นการเปรียบเทียบข้อมูลจากสำมะโนกับข้อมูลทางประชากรค่าสิทธิ์ ซึ่งสามารถทำได้หลายวิธีโดยการคำนวณค่าที่คาดว่าจะนับได้จากสำมะโน (Expected value) แล้วนำค่าที่คำนวณได้นี้มา เปรียบเทียบกับค่าที่นับได้จริง ๆ จากสำมะโน ก็จะทำให้สามารถวัดค่าความคลาดเคลื่อนของคัมรวม ของจำนวนของประชากรทั้งหมด และค่าความคลาดเคลื่อนสุทธิของแต่ละกลุ่ม ซึ่งผลด้วยกัน 2 วิธี คือ

ก. การใช้สมการสมดุลย์ (Balancing equation) วิธีการนี้ต้องอาศัยผลที่ได้จากการทำสำมะโนครั้งก่อน จำนวนคนเกิด คนตาย และจำนวนผู้ที่มีการเคลื่อนย้ายระหว่างเวลาของการทำสำมะโนสองครั้ง ซึ่งจำนวนคนเหล่านี้ได้มาจากการจดทะเบียนหรือบันทึกของทางราชการ แล้วนำจำนวนต่าง ๆ เหล่านี้มาคำนวณหาจำนวนประชากรที่คาดว่าจะนับได้จากการทำสำมะโนครั้ง นี้ ซึ่งสามารถคำนวณได้เฉพาะยอดรวมใหญ่ ๆ เท่านั้น

สูตรของสมการสมดุลย์ คือ

$$P_1 = P_0 + B - D + I - E$$

โดยที่

$P_1$  = จำนวนประชากรที่คาดว่าจะนับได้จากการทำสำมะโนครั้งนี้

$P_0$  = จำนวนประชากรจากสำมะโนครั้งก่อน

B = จำนวนคนเกิดระหว่างเวลาของการทำสำมะโนสองครั้ง

D = จำนวนคนตายระหว่างเวลาของการทำสำมะโนสองครั้ง

I = จำนวนผู้ที่ย้ายเข้าในระหว่างเวลาของการทำสำมะโนสองครั้ง

E = จำนวนผู้ที่ย้ายออกในระหว่างเวลาของการทำสำมะโนสองครั้ง

จากค่าที่คำนวณได้นี้ทำให้สามารถหาค่าของความคลาดเคลื่อนของคู่ผสมสุทธิ (Net coverage errors) ของสำมะโน ซึ่งก็คือผลต่างระหว่างจำนวนประชากรที่กล่าวจะนับ ได้จากการทำสำมะโนครั้งนี้ กับจำนวนประชากรที่นับได้จริงจากสำมะโนครั้งนี้

ข. การศึกษาจากกลุ่มคนที่เกิดรอด (Study of cohort survival) ตัวอย่าง เช่นการเปรียบเทียบจำนวนคนในกลุ่มอายุ 45 ปีของการทำสำมะโนครั้งนี้ กับจำนวนคนในกลุ่มอายุ 35 จากการทำสำมะโนครั้งก่อนเมื่อ 10 ปี ที่ผ่านมาโดยได้มีการปรับตัวเลขด้วยจำนวนคนตายและ จำนวนคนที่มีการเคลื่อนย้าย ซึ่งผลต่างจากการเปรียบเทียบนี้ จะได้ค่าประมาณของความคลาดเคลื่อน สุทธิ (Net error) ของแต่ละกลุ่มอายุ

#### ข้อดีข้อเสียของการวิเคราะห์ทางด้านประชากรศาสตร์

ข้อดีของการวิเคราะห์ทางด้านประชากรศาสตร์ คือ

1. ไม่ต้องอาศัยการเปรียบเทียบ (Matching) ระหว่างข้อมูลจากสอง แหล่ง ซึ่งเป็น การลดความยุ่งยากไปได้อย่างมาก
2. ไม่มีความคลาดเคลื่อนที่เกิดจากการใช้ตัวอย่าง เนื่องจาก การวัดความคลาดเคลื่อนด้วยวิธีนี้จะใช้ข้อมูลทั้งหมดจึงไม่ต้องมีการเลือกตัวอย่าง
3. ในการวัดความคลาดเคลื่อนด้วยวิธีนี้เสียค่าใช้จ่ายน้อยกว่าวิธีอื่น
4. ได้อัตราการตกแฉงนับในระดับหมวดอายุ (ถ้าใช้วิธี เอ.)

ข้อเสียของการวิเคราะห์ทางด้านประชากรศาสตร์ คือ

1. ในการใช้วิธี ก. นั้น วัดความคลาดเคลื่อนได้เฉพาะยอดรวมใหญ่ ๆ คือ วัดได้เฉพาะค่าความคลาดเคลื่อนของคู่ผสมใน ระดับประเทศเท่านั้น และไม่สามารถวัดความคลาดเคลื่อนของการแจ้งลักษณะได้
2. ไม่สามารถทราบได้ว่าความคลาดเคลื่อนของคู่ผสมสุทธิที่คำนวณได้นั้น เป็น ความคลาดเคลื่อนของคู่ผสมสุทธิที่เกิดจากการทำสำมะโนครั้งนี้หรือครั้งก่อน หรือเกิดจากจำนวนคนเกิด คนตาย คนที่เคลื่อนย้าย

3. เนื่องจากค่าที่จะประมาณขึ้นมาจะต้องขึ้นอยู่กับค่าต่าง ๆ ที่ได้จากการจดทะเบียน ดังนั้นถ้าการจดทะเบียนไม่ดีพอ (ซึ่งประเทศส่วนมากก็ยังมีระบบการจดทะเบียนที่ยังไม่สมบูรณ์) ก็จะทำให้ค่าที่ประมาณได้ผิดพลาดไปด้วย

4. ผู้วิเคราะห์จะต้องมีความรู้ทางประชากรศาสตร์

### 3. วิธีการประเมินค่าความคลาดเคลื่อนโดยตรงหรือการประมาณค่าจากสองระบบ

(Direct methods of evaluation หรือ Dual system estimation)

เป็นวิธีการวัดความคลาดเคลื่อนที่ต้องอาศัยการเปรียบเทียบของข้อมูลที่ได้จากสำมะโนกับข้อมูลที่ได้จากแหล่งต่าง ๆ แบบรายการต่อรายการ (Case-by-Case Comparison) ซึ่งมีส่วนด้วยกันหลายวิธี ดังนี้คือ

ก. โดยการสำรวจภายหลังการแจงนับ (Post Enumeration Survey)

ข. การใช้ข้อมูลจากบันทึกต่าง ๆ (Record Check)

ค. การใช้การสำรวจอย่างอื่น ๆ

ก. การสำรวจภายหลังการแจงนับ (Post Enumeration Survey)

ซึ่งอาจจะเรียกว่าการสำรวจโดยการสัมภาษณ์ซ้ำ (Reinterview Survey)

เป็นวิธีที่ใช้กันอย่างแพร่หลายที่สุดใน การวัดความคลาดเคลื่อนของคัมรวม และความคลาดเคลื่อนของการแจ้งลักษณะ การทำสำรวจนี้ต้องทำหลังจากการแจงนับของสำมะโนแล้ว และใช้พนักงานคนละชุดกับสำมะโน เพื่อที่จะได้เป็นอิสระจากสำมะโน แต่ต้องอยู่ภายใต้ข้อกำหนด (Condition) เดียวกันกับสำมะโน ทำโดยวิธีการเลือกท้องที่ตัวอย่างจากรายชื่อเขตสำมะโนเพื่อออกไปทำการนับจุดและแจงนับใหม่ แล้วนำข้อมูลที่ได้จากการทำสำรวจภายหลังการแจงนับนี้มา เปรียบเทียบกับข้อมูลที่ได้จากการทำสำมะโน เพื่อหาอัตราความครบถ้วนของการแจงนับ, อัตราการตกแจงนับ, อัตราความแตกต่างลุ่ม, อัตราการเคลื่อนลุ่ม, ดัชนีของความไม่แน่นอน และดัชนีของความไม่แน่นอนแบบแอล-โพลด์

ในการวัดความคลาดเคลื่อนของคัมรวมจะนำหน่วยแจงนับที่ได้จากท้องที่ตัวอย่างของการสำรวจภายหลังการแจงนับมา เปรียบเทียบกับหน่วยแจงนับของสำมะโน ส่วนการวัดความคลาดเคลื่อน

ของการแจ้งลักษณะและอาศัยข้อมูลจากครัวเรือนหรือคนซึ่งเปรียบเทียบกันได้ในท้องที่ตัวอย่างมาใช้ในการเปรียบเทียบ

### ข้อดีและข้อเสียของการใช้การสำรวจภายหลังการแจงนับ

ข้อดีของการใช้การสำรวจภายหลังการแจงนับ คือ

1. ได้ข้อมูลที่สามารนำมาใช้ในการแยกวัดความคลาดเคลื่อนของคัมรวมและความคลาดเคลื่อนของการแจ้งลักษณะ

2. สามารถแยกวิเคราะห์ส่วนประกอบของความคลาดเคลื่อนของคัมรวมออกเป็น

ก. แจงนับขาด (Under-Enumeration)

ข. แจงนับเกิน (Over-Enumeration)

ค. คนที่ตกแจงนับในบ้านที่ตกแจงนับ

ง. คนที่ตกแจงนับในบ้านที่แจงนับมา

ซึ่งจากลักษณะต่าง ๆ นี้ทำให้สามารถทราบถึงสาเหตุของการเกิดความคลาดเคลื่อนของคัมรวมได้ดีขึ้น เพื่อจะได้นำไปปรับปรุงในสำมะโนครั้งต่อไป

3. สามารถทราบถึงลักษณะของหน่วยที่มักจะตกแจงนับ เช่น บ้านที่ตั้งอยู่ในสถานที่ที่เห็นไม่ค่อยชัด เจนจากถนน หรือบ้านที่ตั้งอยู่ในบริเวณที่มีกิ่งก้านอาศัยไม่ค่อยหนาแน่น จะพบว่าบ้านที่มีลักษณะเช่นนี้มักจะตกแจงนับ ซึ่งจะได้นำไปปรับปรุงแผนการต่าง ๆ ในการทำสำมะโนครั้งต่อไป

4. สามารถเพิ่มคำถามทุกอย่างที่ต้องการใช้ในการเปรียบเทียบกับสำมะโนได้

ข้อเสียของการใช้การสำรวจภายหลังการแจงนับ

1. เสียค่าใช้จ่ายในการทำสำรวจมาก เพราะจะต้องออกไปทำการนับจุดและแจงนับใหม่

2. มีความยุ่งยากมากในการเปรียบเทียบข้อมูลระหว่างการทำสำรวจและสำมะโน ทั้งในการหาความคลาดเคลื่อนของคัมรวม และความคลาดเคลื่อนของการแจ้งลักษณะ ซึ่งจะทำให้



เกิดความผิดพลาดในการเปรียบเทียบ และจะส่งผลทำให้ค่าประมาณของอัตราต่าง ๆ ผิดไปด้วย ก็  
จะส่งผลทำให้ค่าความแปรปรวนของความคลาดเคลื่อนทั้งสองชนิดสูงขึ้นด้วย

3. ในการทำสำรวจภายหลังการแจกแจงนั้นก็มีความคลาดเคลื่อนอยู่ในตัว  
เองด้วย ซึ่งความคลาดเคลื่อนนี้จะมีแนวโน้มที่จะมีความสัมพันธ์กับความคลาดเคลื่อนที่เกิดจากสำมะโน  
คือคนที่ตกแจกแจงในสำมะโนมักจะตกแจกแจงในการสำรวจภายหลังการแจกแจงด้วย เช่นทั้งสำมะโนและ  
การสำรวจภายหลังการแจกแจงมักจะตกคนที่ไม่มีที่อยู่เป็นประจำเหมือนกัน

4. ได้ อัตราการตกแจกแจงในระดับทั่วประเทศหรือภาค ไม่อาจได้ในระดับ  
หมวดอายุ เนื่องจากขนาดตัวอย่างจำกัด

#### ย. ไปข้อมูลจากบันทึกต่าง ๆ (Record Check)

เป็นวิธีที่ใช้วัดความคลาดเคลื่อนโดยการเปรียบเทียบระหว่างข้อมูลจากสำมะโน  
และข้อมูลจากบันทึกต่าง ๆ จากแหล่งอื่นซึ่งมีวิธีการเก็บรวบรวมข้อมูลเป็นอิสระจากสำมะโน ข้อมูลที่  
จะไปใช้ในการเปรียบเทียบนี้ต้องเป็นชนิดเดียวกัน ซึ่งต้องมีมากพอและสมบูรณ์พอที่จะใช้ในการเปรียบเทียบ  
ข้อมูลที่จะนำมาเปรียบเทียบกับสำมะโนนี้อาจได้มาจากบันทึกต่าง ๆ เช่น

ก. รายชื่อของคนที่ถูกแจกแจงจากสำมะโนครั้งก่อน  
ข. ทะเบียนของคนเกิดระหว่างการทำสำมะโนครั้ง  
ค. รายชื่อ นักเรียน, นิสิต หรือนักศึกษาที่เข้าเรียนในโรงเรียน วิทยาลัย หรือ  
มหาวิทยาลัย

ง. รายชื่อผู้มีบัตรประจำตัวประชาชน

จ. รายชื่อผู้มีสิทธิเลือกตั้ง

ฉ. รายชื่อผู้มีใบขับขี่รถยนต์

ช. รายชื่อบุคคลจากทะเบียนราษฎร

การใช้ข้อมูลจากบันทึกต่าง ๆ เพื่อวัดความคลาดเคลื่อนของคู่รวม ข้อมูลที่จะนำ  
มาใช้ในการเปรียบเทียบอาจจะเป็นข้อมูลของทั้งหมดหรือข้อมูลจากตัวอย่างที่เลือกขึ้นมาบางส่วนก็ได้

แต่ส่วนมากจะใช้วิธีเลือกตัวอย่างมาจำนวนหนึ่งจากรายชื่อ (list) ที่จะนำมาใช้เปรียบเทียบกับ สัมมะโนหารายละเอียดเกี่ยวกับที่อยู่อาศัยซึ่งควรจะถูกแจกแจงนับในสัมมะโนกับรายละเอียดเฉพาะอื่น ๆ แล้วนำข้อมูลที่ได้ไปเปรียบเทียบข้อมูลจากสัมมะโน ถ้ารายใดเปรียบเทียบไม่ได้ (Unmatched) ก็ต้องออกไปตรวจสอบในสนามใหม่ (Field follow-up) เพื่อดูว่ามีการตกแจกแจงนับในสัมมะโนจริงหรือไม่ แล้วนำไปหาจำนวนที่เปรียบเทียบได้ ซึ่งผลจากการเปรียบเทียบนี้จะสามารถหาจำนวนที่ตกแจกแจงนับ (Under-Enumeration) ในสัมมะโนได้

การนำข้อมูลจากบันทึกต่าง ๆ มาใช้ในการวัดความคลาดเคลื่อนของการแจง สัมมะโนนั้นจะใช้ลักษณะของบุคคลตัวอย่าง ซึ่งเลือกมาจากรายชื่อของบันทึก แล้วนำไปเปรียบเทียบ กับลักษณะที่ได้จากสัมมะโน ผลจากการเปรียบเทียบเหล่านี้จะนำไปประมาณหาค่าความแปรปรวนของ ค่าตอบ (Response Variance) และความเอนเอียงของค่าตอบ (Response Bias) ในการ เปรียบเทียบนี้มีข้อสมมุติว่าข้อมูลที่ได้จากบันทึกต่าง ๆ มีความถูกต้องมากกว่าข้อมูลจากสัมมะโน เช่น วัน เดือน ปีเกิด สถานที่เกิด จากใบสูติบัตร บ่งชี้ถูกต้องกว่าที่ได้จากสัมมะโน

#### ข้อดีและข้อเสียของการใช้บันทึกต่าง ๆ

ข้อดีของการใช้บันทึกต่าง ๆ

1. บันทึกต่าง ๆ เหล่านี้ เป็นแหล่งข้อมูลที่มีความเป็นอิสระจากสัมมะโนมากกว่า การสำรวจภายหลังการแจกแจง
2. การตกจากบันทึกมีความสัมพันธ์กับการตกจากสัมมะโนน้อยกว่า การตกจากการสำรวจภายหลังการแจกแจง ดังนั้นการใช้ข้อมูลจากบันทึกจึงสามารถนำมาใช้วัดความคลาดเคลื่อนได้ดีกว่า
3. จากการศึกษาเชื่อว่าข้อมูลจากบันทึกต่าง ๆ มีความถูกต้องมากกว่าข้อมูลจากสัมมะโน ก็จะทำให้สามารถหาค่าความคลาดเคลื่อนของการแจง สัมมะโนซึ่งได้จากการ เปรียบเทียบระหว่างข้อมูลจากบันทึกต่าง ๆ กับข้อมูลจากสัมมะโนได้ผลใกล้เคียงความเป็นจริงมากกว่า

ข้อเสียของการใช้บันทึกต่าง ๆ

1. ข้อมูลจากบันทึกจำนวนมากมักจะมียารายการเพียงชนิดเดียว เช่นถ้าจะใช้ข้อมูลจากใบคู่ติบัตร ก็จะได้ข้อมูลเกี่ยวกับการเกิดเท่านั้น ทำให้เปรียบเทียบได้น้อยรายการ
2. การเปรียบเทียบระหว่างข้อมูลจากบันทึกและข้อมูลจากสำมะโนทำได้ยาก เพราะว่าการเรียงลำดับของข้อมูลจากทั้งสองแหล่งผิดกัน
3. ข้อมูลเกี่ยวกับที่อยู่อาศัย เช่น บ้านเลขที่ หรือสถานที่ตั้งของครัวเรือนจากบันทึกไม่ค่อยทันสมัย และไม่ครอบคลุมทั่วหนัก
4. เสียค่าใช้จ่ายในการเก็บรวบรวมข้อมูลเกี่ยวกับรายละเอียดของคนที่จะนำมาเปรียบเทียบสูงมาก โดยเฉพาะถ้าไม่มีบ้านเลขที่หรือชื่อถนน ซึ่งทำให้ค้นหาได้ยากมาก

#### ค. การใช้การสำรวจอย่างอื่น

เป็นการสำรวจที่ไม่ได้ทำขึ้นเพื่อวัดความคลาดเคลื่อนของสำมะโนโดยเฉพาะ แต่เป็นการทำสำรวจโดยทั่ว ๆ ไป เพื่อวัตถุประสงค์ที่ต่าง ๆ กัน เช่นการสำรวจเกี่ยวกับครัวเรือน การสำรวจแรงงาน การสำรวจทางด้านประชากรศาสตร์ ฯลฯ การจะนำการสำรวจเหล่านี้มาใช้ในการวัดความคลาดเคลื่อนของสำมะโนนั้น ก็โดยอาศัยข้อมูลที่ได้จากการทำสำรวจนี้มา เปรียบเทียบกับข้อมูลจากสำมะโน ถ้าข้อมูลที่จะใช้ในการเปรียบเทียบมีรายละเอียดไม่เพียงพอ ก็อาจเพิ่มคำถามเฉพาะในการวัดความคลาดเคลื่อนของสำมะโนเข้าไปในคำถามของการสำรวจนั้น ๆ การทำสำรวจนี้จะทำก่อนหรือหลังการทำสำมะโนก็ได้ โดยไม่สนใจว่าข้อมูลที่ได้จากการสำรวจนี้เป็นข้อมูลในวันสำมะโนหรือไม่ สิ่งจำเป็นต้องมีวิธีการ คำถาม และการแจกแจงเป็นพิเศษ เพื่อที่จะได้ข้อมูลตามที่ต้องการในการเปรียบเทียบกับข้อมูลจากสำมะโน และเพื่อที่จะปรับข้อมูลที่ต่างกันเนื่องจากเวลาในการทำสำรวจและสำมะโนที่ต่างกัน เช่นการรวมแตกต่างกันในเรื่องค่าจ้าง อาชีพ และการเปลี่ยนงานในระหว่างวันที่สำมะโนและวันที่ทำสำรวจ

### ข้อดีและข้อเสียของการใช้การสำรวจอย่างอื่น

#### ข้อดีของการใช้การสำรวจอย่างอื่น

1. การทำสำรวจอย่างอื่นมีความเป็นอิสระจากสภาวะโนมากรกว่า การสำรวจภายหลังการแจกจ่าย แต่น้อยกว่าการใช้ข้อมูลจากบันทึกต่าง ๆ

2. การทำสำรวจอย่างอื่นมีเสถียรค่าใช้ลายน้อยกว่าการทำสำรวจภายหลังการแจกจ่าย เพราะเป็นผลพลอยได้จากการสำรวจประจำ

#### ข้อเสียของการใช้การสำรวจอย่างอื่น

1. ข้อมูลจากการใช้การสำรวจอย่างอื่นมักมีเพียงไม่กี่รายการที่จะสามารถนำมาใช้เปรียบเทียบข้อมูลจากสภาวะโนได้

2. การเปรียบเทียบระหว่างข้อมูลจากการสำรวจอย่างอื่นและข้อมูลจากสภาวะโนทำได้ยาก เพราะการเรียงลำดับของข้อมูลจากทั้งสองแหล่งไม่เหมือนกัน และเนื่องจากเวลาที่เก็บรวบรวมข้อมูลก็ไม่ตรงกัน

#### 4. วิธีทางสถิติ (Statistical Methods)

เป็นอีกวิธีหนึ่งที่จะนำมาใช้ในการวัดค่าความคลาดเคลื่อนของตัวรวม ซึ่งมีวิธีการวัดหลายวิธีด้วยกันคือ

ก. วิธีสังเคราะห์ (Synthetic Methods)

ข. วิธีการประมาณค่าสหสัมพันธ์ (Corelation Estimating Methods)

วิธีทางสถิตินี้ยังเป็นวิธีการที่ใหม่ ดังนั้นเอกสารเกี่ยวกับวิธีการนี้จึงมีน้อย และยังไม่เป็นที่แพร่หลาย และเนื่องจากวิทยานิพนธ์ฉบับนี้ได้สนใจวิธีการนี้ จึงขอไม่บรรยายรายละเอียดเกี่ยวกับเรื่องนี้

จากที่กล่าวมาแล้วนั้นจะพบว่าไม่มีวิธีใดวิธีหนึ่งที่จะเป็นวิธีที่ดีที่สุดในการที่จะวัดความคลาดเคลื่อนของสภาวะโน และวิธีที่ประเทศหนึ่งใช้แล้วได้ผลก็อาจจะใช้ไม่ได้ผลในอีกประเทศหนึ่ง

ก็ได้ แต่วิธีการที่ใช้กันมากที่สุดก็คือ การใช้การสำรวจภายหลังการแจกฉั่ม ดังนั้นในการวัดความคลาดเคลื่อนของฉั่มจะ ไม่ประชากรและเคหะทดลองปี 2522 นี้จะวัด โดยการใช่วิธีการสำรวจภายหลังการแจกฉั่มจึง เป็นวิธีที่นักวิชาการสถิติทำการศึกษาค้นและปรับปรุงแก้ไขข้อบกพร่องของวิธีการเดิมอยู่เสมอ ๆ

## 2.1 วิธีการทำการสำรวจภายหลังการแจกฉั่ม

การสำรวจภายหลังการแจกฉั่มเป็นการสำรวจที่ทำขึ้นหลังจากการทำฉั่มจะ ไม่เสร็จแล้ว โดยจะพยายามทำให้เร็วที่สุดเพื่อป้องกันไม่ให้ผู้ถูกสัมภาษณ์ลืมคำตอบ หรือลืมเหตุการณ์ในวันฉั่มจะ ไม่ และเพื่อไม่ให้มีผู้เคลือบอายมากเกินไป ซึ่งจะเป็นการยุ่งยากต่อการเปรียบเทียบ แต่จะต้องทำการสำรวจภายหลังการแจกฉั่มหลังจากที่พนักงานแจกฉั่มของฉั่มจะ ไม่ทำงานสนามเสร็จเรียบร้อยแล้ว เพื่อให้ข้อมูลที่ได้จากสองแหล่งเป็นอิสระต่อกัน การทำการสำรวจภายหลังการแจกฉั่มต้องพยายามทำให้เป็นอิสระจากการทำฉั่มจะ ไม่แต่ต้องอยู่ภายใต้ข้อกำหนดเดียวกันกับฉั่มจะ ไม่ โดยการเลือกท้องที่ตัวอย่างจากฉั่มจะ ไม่ ระยะเวลาท้องที่ตัวอย่างที่เลือกมาต้องมีขนาดไม่ใหญ่กว่าขนาดของท้องที่ในฉั่มจะ ไม่ เช่นในฉั่มจะ ไม่แบ่งท้องที่ออกเป็น เขตแจกฉั่ม ซึ่งประกอบด้วย เขตคอนหนึ่ง เขตคอนหรือหลายเขตคอน แต่ในการทำการสำรวจภายหลังการแจกฉั่มจะ นำเขตการแจกฉั่มมาแบ่งออกเป็นเขตคอน แล้วจึงเลือกเขตคอนตัวอย่างขึ้นมาทำการสำรวจ โดยออกไปทำการนับจุดและแจกฉั่มใหม่ ซึ่งวิธีการจุดและแจกฉั่มนี้จะทำเหมือนในฉั่มจะ ไม่ โดยพยายามเน้นถึงข้อมูลในวันฉั่มจะ ไม่ แล้วจึงนำข้อมูลที่ได้จากการทำการสำรวจภายหลังการแจกฉั่มมาเปรียบเทียบกับข้อมูลที่ได้จากฉั่มจะ ไม่ นำผลของการเปรียบเทียบนั้นมาคำนวณหาค่าต่าง ๆ ที่จะใช้ในการวัดความครบถ้วนและความถูกต้องของฉั่มจะ ไม่ต่อไป

สรุปขั้นตอนของการสำรวจภายหลังการแจกฉั่มมีดังนี้

1. ให้คำจำกัดความประชากร (universe) ตามฉั่มจะ ไม่
2. เลือกท้องที่ตัวอย่าง โดยอาศัยทฤษฎีการสุ่มตัวอย่าง ซึ่งท้องที่นั้นจะต้องไม่ใหญ่กว่า

เขตแจกฉั่มของฉั่มจะ ไม่

3. ทำงานสัปดาห์ภายหลังงานสามะโน โดยพนักงานคนละชุดกับสามะโน
4. เปรียบเทียบข้อมูลรายครัวเรือนและรายบุคคลของสามะโนและของการสำรวจภายหลัง

#### การแจงนับ

5. ตรวจสอบในรายที่ลงสับ โดยวิธีตัวอย่าง เพื่อหาข้อสรุป
6. ประมาณค่าต่าง ๆ
7. วิเคราะห์ผลที่ได้
8. เขียนรายงานสรุป

วิธีการสำรวจภายหลังการแจงนับเพื่อใช้วัดความคลาดเคลื่อนของข้อมูลรวมมีอยู่ด้วยกัน 3 วิธี <sup>(1)</sup>

1. การสำรวจภายหลังการแจงนับ แบบ A
2. การสำรวจภายหลังการแจงนับ แบบ B
3. การสำรวจภายหลังการแจงนับ แบบ C

แต่ละแบบจะมีวิธีการปฏิบัติอยู่เคลื่อนย้ายและวิธีการวัดอัตราต่าง ๆ แตกต่างกันไป

#### การสำรวจภายหลังการแจงนับแบบ A

เป็นวิธีที่รู้จักกันดีและใช้กันมานานแล้ว การสำรวจภายหลังการแจงนับแบบ A นี้ ควรทำเร็วที่สุด (ไม่ควรเกิน 6 เดือน) หลังจากการทำสามะโน เลือกท้องที่ตัวอย่างจากสามะโนขึ้นมาทำการสำรวจโดยการออกไปนับจุดและแจงนับใหม่ วิธีนี้จะสนใจเฉพาะผู้ที่อยู่ในท้องที่ตัวอย่างในวันสามะโนเท่านั้น คือพนักงานแจงนับของการสำรวจภายหลังการแจงนับจะออกไปทำการแจงนับทุกคนในครัวเรือนในท้องที่ตัวอย่าง โดยจะถามว่าในวันสามะโนนั้น เขาอยู่ในครัวเรือนนี้หรือเปล่า และถ้าได้รับคำตอบว่าไม่ได้อยู่ในครัวเรือนนี้ในวันสามะโน พนักงานแจงนับก็จะถามต่อไปว่าใครที่เคยอยู่ในครัวเรือนนี้มาก่อนและตอนนี้ได้ย้ายไปแล้ว ถ้าผู้ที่มาอยู่ใหม่ไม่ทราบว่าคนเดิมย้ายไปอยู่ไหน พนักงานแจงนับก็อาจจะถามจากเพื่อนบ้าน เพื่อที่จะได้ติดตามไปสอบถามรายละเอียดอื่น ๆ ตามสถานที่ที่ผู้นั้นย้ายไปอยู่ใหม่ ทำให้ได้จำนวนผู้ย้ายออก และได้ข้อมูลเกี่ยวกับผู้ย้ายออก จะเห็นได้ว่าวิธี A ให้พนักงานแจงนับพยายามหารายชื่อของคนที่เคยอยู่ในครัวเรือนนี้เมื่อวันสามะโนพร้อมทั้งข้อมูลอื่น ๆ ซึ่งรวมถึง

<sup>1</sup>U.S. Bureau of the Census, Popstan. A Case Study for the 1980 Censuses of Population and Housing. Part A.(U.S.A., 1978),p. 211

ผู้ย้ายออกไปจากท้องที่ตัวอย่างหนึ่ง วันสามะโนด้วย ในการสำรวจนี้จะใช้พนักงานแจกนับอีกชุดหนึ่งซึ่งมีคุณสมบัติกว่าพนักงานแจกนับของสามะโน วิธีนี้จะเห็นเรื่องคุณภาพของงานสำเนาโดยพยายามทำให้งานสำเนา มีความสมบูรณ์และถูกต้องที่สุด เช่นอัตรพนักงานแจกนับเป็นคู่กัน ใช้แผนที่มีรายละเอียดชัดเจนกว่าของสามะโน และควบคุมงานสำเนาอย่างใกล้ชิดที่สุดซึ่งจะช่วยให้ข้อมูลที่ได้จากการสำรวจนั้นมีความสมบูรณ์และครบถ้วนมากกว่าข้อมูลจากสามะโน โดยถือหลักที่ว่า ต้นรวมของข้อมูลที่ได้จากการสำรวจภายหลังการแจกนับด้วยวิธีนี้มีความสมบูรณ์และถูกต้องมากกว่าข้อมูลจากสามะโน ซึ่งจะเป็นการช่วยลดความเอนเอียงสหสัมพันธ์ (Correlation bias)<sup>(2)</sup> ให้มีค่าเท่ากับ 0

นำข้อมูลของผู้ที่ไม่ได้เคลื่อนย้าย (Non-movers) และของผู้ที่ย้ายออก (Out-movers) หลังวันสามะโนที่ได้จากการสำรวจภายหลังการแจกนับนี้ไปเปรียบเทียบกับข้อมูลของทุกคนในท้องที่นั้นจากสามะโนแบบการเปรียบเทียบรายการต่อรายการ เพื่อว่าข้อมูลไหนที่เปรียบเทียบได้ (Matched) และเปรียบเทียบไม่ได้ (Non-matched) ในทุกกรณีของข้อมูลที่เปรียบเทียบไม่ได้ จะต้องมีการออกสำเนาไปทำการตรวจสอบซ้ำ เพื่อเป็นการตัดสินใจว่าข้อมูลที่ได้จากการสำรวจภายหลังการแจกนับและที่ได้จากสามะโนอันไหนถูกต้องกว่ากัน จะได้ไปปรับแก้จำนวนที่เปรียบเทียบได้ และจำนวนที่เปรียบเทียบไม่ได้ ซึ่งจำนวนที่ได้ไว้จะต้องนำไปใช้ในการประมาณค่าความคลาดเคลื่อนต่าง ๆ

สูตรที่ใช้ในการประมาณ

1. อัตราการตกแจกนับของครัวเรือนหรือประชากรในสามะโน ( $s_1$ )

$$s_1 = \frac{U_2}{\hat{N}} \times 100\%$$

2. ยอดรวมของจำนวนครัวเรือนหรือประชากรในสามะโน ( $\hat{N}$ )

$$\hat{N} = N_1 + U_2$$

ในเมื่อ

$N_1$  = ยอดรวมของครัวเรือนหรือประชากรที่แจกนับได้จากสามะโน

$U_2$  = ยอดรวมของครัวเรือนหรือประชากรที่พบในสำรวจภายหลังการแจกนับแต่ไม่พบในสามะโน

<sup>2</sup> ความเอนเอียงสหสัมพันธ์คือความสัมพันธ์ของค่าตัดขอบของการสำรวจภายหลังการแจกนับที่จะมีความเอนเอียงเข้าใกล้ค่าตัดขอบของสามะโน ถ้าความเอนเอียงสหสัมพันธ์มีค่า = 0 แสดงว่าค่าตัดขอบของการสำรวจภายหลังการแจกนับและค่าตัดขอบของสามะโนไม่มีความสัมพันธ์กันเลย

จะเห็นได้ว่าสูตรการประมาณอัตราการตกแรงแบบนี้และการประมาณยอดรวมก็เป็นสูตรที่มีความเอนเอียง เพราะวิธีนี้ตั้งข้อสมมุติฐานว่า การสำรวจภายหลังการแรงแบบสมบูรณ์ที่สุด ดังนั้นจึงไม่มีการตกแรงแบบในการสำรวจภายหลังการแรงแบบไปแต่ค่าที่การสำรวจภายหลังการแรงแบบพบแต่สำมะโนไม่พบเท่านั้น ซึ่งในทางปฏิบัติจะไม่เป็นเช่นนั้น เพราะการสำรวจภายหลังการแรงแบบเองก็มีการตกแรงแบบแต่สำมะโนพบ และยังมีอีกส่วนหนึ่งซึ่งตกทั้งในสำมะโนและการสำรวจภายหลังการแรงแบบ

สูตรที่ควรจะใช้ คือ วิธีประมาณแบบล่องระบบ ดังนี้คือ

1. อัตราความครบถ้วนของการแรงแบบ ของสำมะโน

$$W_1 = \frac{M}{N_2} \times 100\%$$

ในเมื่อ  $M$  = จำนวนที่เปรียบเทียบกับได้

$N_2$  = ยอดรวมที่พบในการสำรวจภายหลังการแรงแบบ

หรืออัตราการตกแรงแบบของสำมะโน ( $S_1$ ) =  $(1 - W_1) \times 100\%$

2. ยอดรวมของจำนวนครัวเรือนหรือประชากรของสำมะโน ( $\hat{N}$ )

$$N = \frac{\hat{N}_1}{W_1}$$

ในเมื่อ  $N_1$  = ยอดรวมที่แรงแบบได้จากสำมะโน

$W_1$  = อัตราความครบถ้วนของการแรงแบบ

ตามประวัติการทำสำรวจภายหลังการแรงแบบของประเทศสหรัฐอเมริกาที่ยังคงใช้วิธี A แบบเก่า เมื่อ ค.ศ. 1950 แต่ได้แก้ไขเปลี่ยนแปลงวิธีประมาณใหม่เมื่อปี 1960 โดยรวมจำนวนที่ตกทั้ง 2 ระบบไว้ในค่าประมาณยอดรวมด้วย ซึ่งก็ยังขาดส่วนที่การสำรวจภายหลังการแรงแบบตกแต่สำมะโนพบอยู่อีก จึงหันมาใช้วิธีประมาณแบบล่องระบบเมื่อ ค.ศ. 1970 คือสำมะโนครั้งที่แล้วนี้เอง

ข้อดีและข้อเสียของการสำรวจภายหลังการแรงแบบแบบ A

ข้อดีของการสำรวจภายหลังการแรงแบบแบบ A คือ

1. เปรียบเทียบข้อมูลระหว่างการทำสำรวจภายหลังการแรงแบบและสำมะโนได้ง่ายเพราะใช้แต่แบบนับคนและแรงแบบของสำมะโนเฉพาะในท้องที่ตัวอย่างเท่านั้น



### ข้อเสียของการทำสำรวจภายหลังการแจกจ่ายแบบ A

1. ทำงานสนามยาก เพราะต้องติดตามไปตามรายละเอียดของผู้ย้ายออกทุกคนไม่ว่าจะย้ายไปไหน
2. รายละเอียดเกี่ยวกับผู้ย้ายออกหลังวันสำมะโนมักได้ไม่ครบ โดยเฉพาะกรณีที่ย้ายออกทั้งครัวเรือน เพราะผู้ที่มาอยู่ใหม่ หรือผู้อื่น (เพื่อนบ้าน) ไม่สามารถจะให้รายละเอียดได้ครบถ้วน ส่งผลให้จำนวนคนย้ายออกต่ำกว่าที่เป็นจริง
3. เสียค่าใช้จ่ายมากในการออกสนามไปทำการตรวจล่อช้ำ เพราะต้องทำทุกกรณีที่เปรียบเทียบไม่ได้ และต้องลงทุนทุกอย่างในการพยายามให้คุณภาพของการสำรวจภายหลังการแจกจ่ายดีกว่าของสำมะโน
4. ค่าประมาณจำนวนประชากรและอัตราการตกแฉงนับที่ได้จากการสำรวจภายหลังการแจกจ่ายนี้จะต่ำกว่าที่เป็นจริงหรืออัตราการตกแฉงนับที่ต่ำกว่าที่ควรจะเป็น การที่อัตราการตกแฉงนับต่ำก็เพราะว่าวิธีการเปรียบเทียบง่ายกว่าและมีการออกสนามไปทำการตรวจล่อช้ำ จึงทำให้ได้ข้อมูลที่เปรียบเทียบกันได้มากขึ้นส่วนหนึ่ง และอีกสาเหตุหนึ่งก็เนื่องมาจากสูตรที่ใช้ในการประมาณค่าประชากร ( $\hat{N}$ )

$$\hat{N} = N_1 + U_2$$

ยังขาดส่วนประกอบสำคัญอีกสองส่วน คือที่สำมะโนพบ แต่การสำรวจภายหลังการแจกจ่ายไม่พบ และที่ตกทั้งสำมะโนและการสำรวจภายหลังการแจกจ่ายตามที่ได้กล่าวแล้ว

### 5. มีความเอนเอียงสหสัมพันธ์ (Correlation bias) สูง

#### การสำรวจภายหลังการแจกจ่ายแบบ B

การสำรวจภายหลังการแจกจ่ายแบบ B ค้นพบโดย Dr. Eli S. Marks ซึ่งเป็นที่ปรึกษาของสำนักงานสถิติของสหรัฐอเมริกา (U.S. Bureau of the Census) และเป็นผู้เชี่ยวชาญในเรื่องการประเมินผลการทำสำมะโนประชากร ในปี พ.ศ. 2517 ได้มีประเทศต่าง ๆ ได้ทดลองใช้วิธีนี้ เช่น ประเทศเกาหลี ปารากวัย และ โดมินิกัน ส่วนสหรัฐอเมริกาเองได้ทดลองใช้ในสำมะโนทดลองเมื่อ พ.ศ. 2521

หลักการโดยทั่วไปของการสำรวจภายหลังการแจกแจงแบบ B ก็คือ เลือกห้องที่ตัวอย่าง จากสำมะโนขึ้นมาทำการนับจุดและแจกแจงใหม่ จะสนใจ เฉพาะคนที่อยู่ในห้องที่ตัวอย่าง ในขณะที่ทำการสำรวจภายหลังการแจกแจงเก็บเท่านั้น แต่จะถามถึงสถานที่ที่ตรวจอยู่ในวันสำมะโน โดยจะถามว่าในเวลา ทำสำมะโนเขาอาศัยอยู่ในครัวเรือนนี้หรือเปล่า ถ้าไม่อยู่ก็จะถามว่าย้ายมาจากที่ไหน ซึ่งจะเป็นประโยชน์ในการชักแบบจากสำมะโนเพื่อนำมาใช้ในการเปรียบเทียบข้อมูล จะทำให้ได้จำนวนผู้ย้ายเข้า หลังวันสำมะโน การสำรวจวิธีนี้ควรทำหลังการทำสำมะโนไม่เกิน 12 เดือน การสำรวจด้วยวิธีนี้ ไม่จำเป็นต้องใช้พนักงานแจกแจงที่มีคุณภาพดีกว่าพนักงานที่ทำในสำมะโน อาศัยพนักงานที่มีคุณภาพพอ ๆ กันก็ได้ แต่วิธีนี้จะเน้นเรื่องความเป็นอิสระซึ่งกันและกันของการทำสำรวจภายหลังการแจกแจงและสำมะโน โดยจะถือหลักที่ว่าความคลาดเคลื่อนของการทำสำรวจภายหลังการแจกแจงและสำมะโนเกิดขึ้นอย่าง เป็นอิสระซึ่งกันและกัน แล้วก็จะทำให้ค่าความเอนเอียงสัมพันธ์มีค่าเท่ากับ 0 จะเห็นว่าวิธี B นี้ แจกแจงคนที่อยู่ในขณะทำสำรวจภายหลังการแจกแจงซึ่งรวมทั้งคนที่ย้ายเข้า หลังวันสำมะโนด้วย

นำข้อมูลของผู้ไม่ได้เคลื่อนย้าย และของผู้ย้ายเข้าที่ได้จากการทำสำรวจภายหลังการแจกแจง มาเปรียบเทียบกับข้อมูลของสำมะโนและในกรณีของผู้ย้ายเข้าก็จะชักแบบสำมะโนจากห้องที่อื่นที่ผู้ย้ายเข้าตอบว่า เคยอยู่มาก่อน ในการเปรียบเทียบจะใช้วิธีเปรียบเทียบทางเดียว (One - way matching) โดยถือข้อมูลจากการสำรวจภายหลังการแจกแจงเป็นหลัก ในกรณีที่เปรียบเทียบไม่ได้ จะพยายามให้หลักการออกส่นามไปทำการตรวจล่องบ้านน้อยที่สุด ผลจากการเปรียบเทียบจะนำไปคำนวณหา อัตราความครบถ้วนของการแจกแจงและอัตราการตกแจกแจงได้

สูตรที่ใช้ในการประมาณ

สูตรในการประมาณค่าต่าง ๆ จะใช้วิธีการประมาณผลจากข้อมูลแบบสองระบบ

PES	สำมะโน		
	พบ	ไม่พบ	รวม
พบ	M	$U_2$	$N_2$
ไม่พบ	$U_1$	Z	$V_2$
รวม	$N_1$	$V_1$	N

โดยที่

$M$  = จำนวนที่เปรียบเทียบได้

$N_1$  = จำนวนทั้งหมดที่พบในสำมะโน

$N_2$  = จำนวนทั้งหมดที่พบในการสำรวจภายหลังการแจกจ่าย

$U_1$  = จำนวนที่พบในสำมะโนแต่ไม่พบในการสำรวจภายหลังการแจกจ่าย

$U_2$  = จำนวนที่พบในการสำรวจภายหลังการแจกจ่ายแต่ไม่พบในสำมะโน

แต่เนื่องจากค่า  $N_2$  และ  $M$  เป็นค่าที่เราประมาณขึ้นมาจากตัวอย่าง จึงจะใช้  $\hat{N}_2$  และ

$\hat{M}$  แทน

1. อัตราความครบถ้วนของการแจกจ่ายของสำมะโน ( $w$ )

$$w = \frac{\hat{M}}{\hat{N}_2} \times 100 \%$$

$$\text{อัตราการตกแจกจ่ายของสำมะโน} = (1-w) \times 100 \%$$

2. ยอดรวมของจำนวนครัวเรือนหรือประชากร ของสำมะโน ( $\hat{N}$ )

$$\begin{aligned} \hat{N} &= \frac{N_1}{w} \\ &= \frac{N_1 \hat{N}_2}{\hat{M}} \end{aligned}$$

### ข้อดีและข้อเสียของการทำสำรวจภายหลังการแจกจ่ายแบบ B

ข้อดีของการสำรวจภายหลังการแจกจ่ายแบบ B คือ

1. งานสนามง่ายกว่าการสำรวจภายหลังการแจกจ่ายแบบ A เพราะถามเฉพาะคนในท้องที่ตัวอย่างเท่านั้น ไม่ต้องติดตามไปตามข้อมูลของผู้ย้ายออก ไม่มีข้อยุ่งยากในการเก็บข้อมูลของผู้ย้ายออก

2. ได้ข้อมูลเกี่ยวกับผู้ย้ายเข้าค่อนข้างครบถ้วนและถูกต้อง เพราะเป็นการถามจากผู้ย้ายเข้าเอง หรือสัมภาษณ์ในครัวเรือนของผู้ย้ายเข้า

3. วิธีการประมาณค่าต่าง ๆ นั้น จะได้ค่าประมาณที่ไม่มีความเอนเอียง เพราะใช้วิธีการประมาณจากข้อมูลสองระบบ ซึ่งได้ค่าถึงถึงข้อมูลที่ไม่ว่างทั้งในการสำรวจภายหลังและสำมะโนเข้าไปด้วยแล้ว

ข้อเสียของการสำรวจภายหลังการแจงนับแบบ B คือ

1. ในการเปรียบเทียบระหว่างข้อมูลที่ได้จากการสำรวจภายหลังการแจงนับแบบ B กับข้อมูลที่ได้จากสำมะโนนั้นยุ่งยาก และเสียเวลามากกว่าแบบ A เพราะต้องไปชักแบบสำมะโนตามสถานที่ที่อยู่เดิมของผู้ย้ายเข้า ซึ่งอยู่ในท้องที่อื่น และเพิ่มงานในชั้นการแปลงค่าสถิตินำ ให้เป็นรหัสของท้องที่สำมะโน

2. เนื่องจากการสำรวจภายหลังการแจงนับนี้จะทำหลังการทำสำมะโนในระยะห่างพอสมควร จึงทำให้เกิดปัญหาเกี่ยวกับคนเคลื่อนย้ายซึ่งเพิ่มความยุ่งยากในการนำข้อมูลมาเปรียบเทียบกัน

3. ค่าประมาณยอดรวมและอัตราการตกแจงนับจะสูงกว่าที่เป็นจริง หรืออัตราการแจงนับต่ำกว่าที่ควรจะเป็น เนื่องจากวิธีในการเปรียบเทียบข้อมูลยุ่งยากมาก จึงทำให้เกิดความคลาดเคลื่อนในการเปรียบเทียบ (Matching Errors) สูง ทำให้ได้ข้อมูลที่เปรียบเทียบกันได้มีจำนวนน้อย ซึ่งก็จะเกิดเหตุทำให้ค่าอัตราการตกแจงนับสูง

การที่ได้ค่าประมาณยอดรวมสูง ก็เพราะสูตรที่ใช้ในการคำนวณ คือ

$$\hat{N} = \frac{N_1 \hat{N}_2}{\hat{M}}$$

แต่เนื่องจาก  $\hat{M}$  ซึ่งเป็นจำนวนที่เปรียบเทียบกันได้มีค่าต่ำ ก็จะเป็นสาเหตุที่ทำให้ค่าประมาณยอดรวม ( $\hat{N}$ ) มีค่าสูง

4. มีความเอนเอียงในการเปรียบเทียบ (Matching Bias) สูง

การสำรวจภายหลังการแจงนับแบบ C

วิธีนี้เพิ่งค้นพบโดย สำนักงานสถิติของสหรัฐอเมริกาอีกเช่นกัน เมื่อปี 2522 นี้เอง เพราะ

วิธี B ก็ยังมีข้อเสียในเรื่องความยุ่งยากลำบากในการเปรียบเทียบ จึงต้องค้นหาวิธีการใหม่ ๆ ต่อไปอีก วิธี C เป็นการนำวิธีการของการสำรวจภายหลังการแจกนับแบบ A และ แบบ B มารวมเข้าด้วยกัน โดยเอาส่วนดีของแบบ A ในเรื่องการเปรียบเทียบ ซึ่งสะดวกและรวดเร็วกว่ามาใช้กับส่วนดีของแบบ B ในเรื่องงานต้นทางซึ่งตรงไปตรงมา

การสำรวจภายหลังการแจกนับแบบ C นี้ ควรทำหลังการทำสำมะโนแล้วประมาณ 2-3 เดือน วิธีการมีดังนี้ คือ เลือกห้องที่ตัวอย่างจากสำมะโนขึ้นมาทำการสำรวจ โดยจะทำกาแจกนับและแจกนับทุกคนที่อยู่ในห้องที่ตัวอย่างขณะ เวลาทำการสำรวจภายหลังการแจกนับ ซึ่งเป็นวิธีการของแบบ B โดยจะถามทุกคนว่าอยู่ในครัวเรือนนี้มานานเท่าไร ก่อนหรือหลัง วันสำมะโน ถ้าอยู่หลังวันสำมะโนก็จะได้อายุของคนย้ายเข้า และถามทุกคนว่าเมื่อมีผู้ใดย้ายออกหลัง วันสำมะโนบ้าง เพื่อที่จะได้อายุของคนย้ายออก ในกรณีที่ย้ายเข้ามาใหม่ทั้ง ครัวเรือนหลัง วันสำมะโนก็จะถามว่ามีใครบ้างที่อยู่ที่นี่มาก่อนวันสำมะโนรวมทั้งถามรายละเอียดบางอย่างที่เกี่ยวข้องกับผู้ย้ายออก ถ้าผู้ที่ย้ายมาอยู่ใหม่ไม่สามารถตอบคำถามเกี่ยวกับผู้ย้ายออกไปแล้วได้ ก็อาจจะถามจากเพื่อนบ้านก็ได้ ทำให้ได้อายุของผู้ที่ย้ายออกไปแล้ว ซึ่งเป็นวิธีการตามแบบวิธี A ในกรณีบ้านว่าง ถูกรื้อถอน ไฟไหม้หรือย้ายบ้านหลัง วันสำมะโน จะถามรายละเอียดของผู้ที่เคยอยู่ในครัวเรือนนั้น ๆ มาก่อนจากเพื่อนบ้าน ซึ่งก็จะได้อายุของผู้ย้ายออก ในการทำสำรวจภายหลังการแจกนับแบบ C นี้ จะได้อายุของทั้งผู้ย้ายเข้า ผู้ย้ายออก และผู้ที่ไม่ได้เคลื่อนย้าย

นำข้อมูลของผู้ไม่ได้เคลื่อนย้าย และข้อมูลของผู้ย้ายออกที่ได้จากการทำสำรวจภายหลังการแจกนับมา เปรียบเทียบกับข้อมูลของสำมะโนเฉพาะในห้องที่ตัวอย่าง ทำให้ได้จำนวนผู้ที่ไม่ได้เคลื่อนย้ายและผู้ที่ย้ายออกที่เปรียบเทียบได้และสามารถนำไปคำนวณหาสัดส่วนของผู้ย้ายออกที่เปรียบเทียบได้ นำสัดส่วนการ เปรียบเทียบได้นี้ไปประยุกต์กับจำนวนผู้ย้ายเข้า เพื่อหาจำนวนผู้ย้ายเข้าที่เปรียบเทียบได้ เพราะในการสำรวจแบบ C นี้จะถือว่าสัดส่วนการแจกนับได้ของผู้ย้ายเข้า เท่ากับสัดส่วนการแจกนับได้ของผู้ย้ายออก นำจำนวนผู้ที่ไม่ได้เคลื่อนย้ายที่เปรียบเทียบได้รวมกับจำนวนผู้ย้ายเข้าที่เปรียบเทียบได้ก็จะได้จำนวนคนหรือครัวเรือนที่มีอยู่ในที่ตัวอย่างในขณะทำการสำรวจภายหลังการแจกนับ ที่เปรียบเทียบได้ทั้งหมด แล้วนำไปคำนวณหา อัตราความครบถ้วนของการแจกนับและอัตราการตกล้างของสำมะโน

ในการประมาณค่าต่าง ๆ คือยอดรวม และอัตราความครบถ้วนของการแจงนับของการสำรวจแบบ C นี้ เหมือนกับแบบ B คือใช้วิธีการประมาณผลจากข้อมูลแบบสองระบบ

$$1. w = \frac{\hat{M}}{\hat{N}_2}$$

$$2. \hat{N} = \frac{N_1 \hat{N}_2}{\hat{M}}$$

$$= \frac{N_1}{w}$$

#### ข้อดีและข้อเสียของการสำรวจภายหลังการแจงนับแบบ C

ข้อดีของการสำรวจภายหลังการแจงนับแบบ C คือ

1. งานล้นมาง่ายเพราะถามเฉพาะคนในท้องที่ตัวอย่างเท่านั้นซึ่งเป็นการนำส่วนดีของแบบ B มาใช้
2. การเปรียบเทียบข้อมูลระหว่างการทำสำรวจภายหลังการแจงนับและสำมะโนทำได้ง่ายและเสียค่าใช้จ่ายในการเปรียบเทียบน้อยกว่าแบบ B เพราะจะเปรียบเทียบกันเฉพาะในท้องที่ตัวอย่างเท่านั้น ซึ่งเป็นการเปรียบเทียบตามแบบ A
3. วิธีประมาณเป็นวิธีที่ไม่มีการเอนเอียง

ข้อเสียของการสำรวจภายหลังการแจงนับแบบ C คือ

1. เสียค่าใช้จ่ายมาก เพราะจะต้องทำการนับจุดและแจงนับทั้งผู้ย้ายเข้า ผู้ย้ายออก และผู้ไม่ย้าย

#### สรุปผลของการทำสำรวจภายหลังการแจงนับแบบ C<sup>(3)</sup>

1. ให้ค่าประมาณของอัตราความครบถ้วนของแจงนับมากกว่าแบบ B แต่น้อยกว่าแบบ A

<sup>3</sup> Bureau of the Census, Popstan. A case Study for the 1980 Censuses of Population and Housing. Part A (U.S.A., 1978), p. 216

2. มีความเอนเอียงต่ำสัมพัทธ์น้อยกว่าแบบ A
3. มีความเอนเอียงในการเปรียบเทียบน้อยกว่าแบบ B
4. มีความเอนเอียงรวม (Overall biased) น้อยกว่าทั้งแบบ A และแบบ B
5. มีความแปรปรวน (Variance) มากกว่าแบบ B แต่น้อยกว่าแบบ A

เปรียบเทียบวิธีการสำรวจภายหลังจากการแจ้งนับแบบ A, B และ C

แบบ A	แบบ B	แบบ C
1. ทำงานสนามเร็วที่สุดหลังจากสามะโน อย่างช้าไม่เกิน 6 เดือน	1. งานสนามไม่ควรห่างจากสามะโนเกินกว่า 12-18 เดือน	1. งานสนามไม่ควรห่างจากสามะโน มักจะ ทำหลังสามะโนประมาณ 2-3 เดือน
2. ใช้พนักงานที่มีคุณภาพดีกว่าพนักงานที่ทำ สามะโน	2. ใช้พนักงานที่มีคุณภาพพอ ๆ กับพนักงานที่ทำ สามะโน	2. ใช้พนักงานที่มีคุณภาพพอ ๆ กับพนักงานที่ ทำสามะโน
3. เน้นเรื่องคุณภาพงานสนาม พยายามให้ สมบูรณ์และถูกต้องที่สุด เช่น อบรมพนักงาน เป็นพิเศษ ใช้แผนที่มีรายละเอียดและซึก เงินกว่าของสามะโน ควบคุมงานสนาม อย่างใกล้ชิดที่สุด	3. เน้นเรื่องความเป็นอิสระในการทำสำรวจ ภายหลังจากการแจ้งนับ	3. เน้นเรื่องความเป็นอิสระเหมือนแบบ B
4. สนใจเฉพาะคนที่เคยอยู่ในห้องที่ตัวอย่าง เมื่อวันสามะโน ใ้คนักพวกไม่ย้าย และ ย้ายออกหลังวันสามะโน	4. สนใจเฉพาะคนที่อยู่ในขณะทำสำรวจภาย หลังจากการแจ้งนับ ได้แก่พวก ไม่ย้าย และ พวกย้ายเข้าหลังวันสามะโน	4. ถามทุกคนที่อยู่ในครัวเรือนนี้ในขณะที่ทำสำรวจ ภายหลังจากการแจ้งนับ และของผู้ที่ย้ายออก หลังวันสามะโน คือถามทั้งผู้ไม่ย้าย ผู้ย้าย เข้าและผู้ย้ายออก
5. ติดตามไปสอบถามรายละเอียดเกี่ยวกับผู้ ที่ย้ายออกไปแล้ว	5. ไม่ต้องติดตามไปแจ้งนับผู้ย้ายออกไปแล้ว	5. ติดตามสอบถามรายละเอียดเกี่ยวกับผู้ ที่ย้ายออกไปแล้วจากเพื่อนบ้าน
6. เปรียบเทียบกับแบบสามะโนเฉพาะในห้อง ที่ตัวอย่างเท่านั้น	6. เปรียบเทียบกับแบบสามะโนในห้องที่ตัวอย่าง และห้องอื่นที่ผู้ย้ายเข้า เคยอยู่ในเวลา สามะโน	6. เปรียบเทียบคนย้ายออกและไม่ย้าย ในห้อง ที่ตัวอย่างกับห้องของสามะโนเหมือน แบบ A
7. ทำการออกสนามเพื่อตรวจสอบซ้ำทุกกรณี เปรียบเทียบไม่ได้	7. ทำการออกสนามเพื่อตรวจสอบซ้ำน้อยที่สุด	7. ทำการออกสนามเพื่อตรวจสอบซ้ำน้อยที่สุด เหมือนแบบ B
8. ได้จำนวนผู้ย้ายออก และผู้ไม่ได้ย้าย	8. ได้จำนวนผู้ย้ายเข้า และผู้ไม่ได้ย้าย	8. ได้จำนวนผู้ย้ายเข้า ผู้ย้ายออก และผู้ไม่ ได้ย้าย





จากที่กล่าวมาแล้ว จะพบว่าในการทำสำมะโนครัวหลังการแจงนับทุกวิธีต้องอาศัยการเปรียบเทียบข้อมูลที่ได้จากการทำสำมะโนครัวหลังการแจงนับกับ ข้อมูลจากการทำสำมะโนทั้งสิ้น ดังนั้น วิธีการของการเปรียบเทียบจึงเป็นส่วนที่สำคัญมากอย่างหนึ่ง ซึ่งสามารถแบ่งออกเป็นขั้นตอนได้ดังนี้ คือ

1. ค้นหาท้องที่ (Geographic Search) เป็นการค้นหาท้องที่เขตแจงนับของสำมะโนที่จะต้องทำการเปรียบเทียบข้อมูลสำมะโน การทำสำมะโนครัวหลังการแจงนับ ในกรณีของการสำมะโนครัวหลังการแจงนับวิธี A และ C เขตแจงนับของสำมะโนที่ใช้เปรียบเทียบ คือเขตแจงนับซึ่งเขตตอนตัวอย่างของการสำมะโนครัวหลังการแจงนับตั้งอยู่ แต่ถ้าเป็นแบบ B จะต้องค้นหาเขตแจงนับอื่นซึ่งมีผู้รายงานในการสำมะโนครัวหลังการแจงนับว่า เขาได้อยู่เมื่อวันสำมะโน

2. ค้นหาชื่อ (Name Search) ภายในท้องที่

ก. ชั้นที่หนึ่ง (Initial Search) คือดูบ้านเลขที่และรายชื่อหัวหน้าครัวเรือนในเขตตอนตัวอย่างจากการสำมะโนครัวหลังการแจงนับที่ครัวเรือนว่าตรงกับที่ได้ในเขตแจงนับจากสำมะโนหรือไม่ ถ้าหาไม่พบก็จะหาจากเขตแจงนับอื่นที่มีพื้นที่ติดกับเขตตอนตัวอย่าง ถ้ารายชื่อหัวหน้าตรงกันก็จะถือว่าเป็นครัวเรือนที่เปรียบเทียบได้ แล้วจึงเปรียบเทียบรายละเอียดของสมาชิกในครัวเรือนนั้น ๆ เพื่อหาคนที่เปรียบเทียบได้

ข. ชั้นที่สอง (Second Stage Search) เป็นการทำในกรณีที่ครัวเรือนเปรียบเทียบไม่ได้ ซึ่งอาจจะเกิดจากการขอกำลังคนที่ผิดหรือไม่ละเอียดของ ลกคชื่อผิดหรือเพี้ยนไป ในขั้นนี้ไม่เพียงแต่ดูเฉพาะชื่อหัวหน้าครัวเรือนเท่านั้น เพราะชื่อหัวหน้าครัวเรือนอาจจะไม่ตรงกัน เนื่องจากผู้ตอบคนละคนกัน ก็ดูต่อไปว่ามีรายชื่อนั้น ๆ อยู่ในรายชื่อสมาชิกของครัวเรือนในสำมะโนหรือไม่ ถ้าอยู่ที่ถือว่า เป็นครัวเรือนที่เปรียบเทียบได้

3. ตัดสินว่าบุคคลใดเปรียบเทียบได้ ขั้นนี้จะขึ้นอยู่กับกฎการเปรียบเทียบ (Matching rule) คือการตั้งหลักเกณฑ์ว่าควรใช้รายละเอียดอะไรบ้างในการเปรียบเทียบ แต่โดยมากมักใช้ชื่อ เพศ อายุ ความสัมพันธ์กับหัวหน้าครัวเรือน สถานภาพสมรส แล้ววางหลักเกณฑ์ต่อไปว่าในแต่ละรายการนี้จะยอมให้มีผิดพลาดได้แค่ไหนจึงจะถือว่า เปรียบเทียบได้ เช่น เพศไม่ควรผิด

อายุขยมาให้ผิดได้ก็ย เ็นต้น

จากการนำข้อมูลของการทำสำรวจภายหลังการแจกใบมา เปรียบเทียบกับข้อมูลของสำมะโน นั้นทำให้สามารถแยกผลการเปรียบเทียบออกได้ดังนี้ คือ

1. พวกที่เปรียบเทียบได้อย่างชัดเจน (Clearly matched)
2. พวกที่อาจจะเปรียบเทียบได้ (Probably matched)
3. พวกที่อาจจะเปรียบเทียบไม่ได้ (Probably unmatched)
4. พวกที่เปรียบเทียบไม่ได้แน่นอน (Clearly unmatched)

ในกรณีของพวกที่อาจจะเปรียบเทียบได้ และพวกที่อาจจะเปรียบเทียบไม่ได้ จะต้องมีการตัดสินใจกันใหม่ว่าควรจะให้เ็นพวกที่เปรียบเทียบได้แน่นอน หรือเปรียบเทียบไม่ได้แน่นอน ซึ่งอาจจะทำการตัดสินใจกันในสำนักงาน โดยผู้ควบคุมงานทำการตรวจสอบใหม่ หรือต้องเปลี่ยนแปลงกฎการเปรียบเทียบ หรืออาจจะต้องมีการออกสำเนาเพื่อตรวจสอบซ้ำ แต่ในกรณีข้อมูลเปรียบเทียบไม่ได้แน่นอน ถ้าเป็นวิธี A จะต้องมีการออกสำเนาเพื่อตรวจสอบซ้ำทุกราย แต่ถ้าเป็นวิธี B หรือ C ใช้เลือกตัวอย่างในท้องที่ที่เปรียบเทียบไม่ได้มาก ๆ เช่นเกิน 20% ออกไปตรวจสอบใหม่

วัตถุประสงค์ของการออกสำเนาเพื่อตรวจสอบซ้ำ คือ

1. เพื่อใช้เป็นวิธีควบคุมคุณภาพ (Quality Control)
2. เพื่อนำมาใช้ปรับปรุงการเก็บข้อมูลและการเปรียบเทียบ
3. เพื่อเป็นวิธีป้องกันและแก้ไขความคลาดเคลื่อนที่เกิดจากการเปรียบเทียบ

วิธีการทำก็คือจะออกไปทำงานสำเนาใหม่ในกรณีที่เปรียบเทียบไม่ได้ หรือที่เกิดสงสัย โดยจะออกไปตรวจสอบเป็นบางรายการที่ยังสงสัยหรือที่ต้องการตรวจสอบเท่านั้น ไม่ต้องไปตรวจสอบทุกรายการตามที่มีแบบสำรวจภายหลังการแจกใบ เพื่อนำมาใช้ในการตัดสินใจว่าสาเหตุที่เปรียบเทียบไม่ได้นั้นเป็นเพราะอะไร สุ่มควรจะนำมารวมอยู่ในพวกที่เปรียบเทียบได้อย่างชัดเจน หรือพวกที่เปรียบเทียบไม่ได้แน่นอน

จากผลการเปรียบเทียบทำให้สามารถแยกคนออกได้เป็น 3 จำพวก คือ

1. ผู้ที่ไม่ได้เคลื่อนย้าย คือผู้ที่อยู่ในครัวเรือนนั้น ๆ ทั้งในเวลาทำสำมะโนและเวลาทำสำรจภายหลังการแจงนับ
2. ผู้ที่ย้ายออก คือผู้ที่ในเวลาทำสำมะโนอยู่ในครัวเรือนนี้ แต่เวลาทำสำรจภายหลังการแจงนับไม่อยู่เพราะได้ย้ายออกไปแล้ว
3. ผู้ที่ย้ายเข้า คือผู้ที่ในเวลาทำสำมะโนไม่ให้อยู่ในครัวเรือนนี้ แต่เวลาที่สำรจภายหลังการแจงนับอยู่ เพราะเพิ่งย้ายเข้ามา

## 2.2 วิธีการวัดค่าความคลาดเคลื่อนของข้อมูล

ความคลาดเคลื่อนที่สนใจในการวัดมีอยู่ 2 ชนิด คือ ความคลาดเคลื่อนของคุ่มรวม และความคลาดเคลื่อนของการแจ้งลักษณะ ฉะนั้นวิธีการสำรจภายหลังการแจงนับจะเป็นวิธีที่เหมาะสมที่สุด ส่วนวิธีวัดความคลาดเคลื่อนนี้ก็จะแตกต่างกันออกไปดังนี้

### การวัดความคลาดเคลื่อนของคุ่มรวม (Coverage Evaluation)

วัตถุประสงค์ของการวัดความคลาดเคลื่อนของคุ่มรวมคือ

1. เพื่อหาอัตราความครบถ้วนของการแจงนับ และอัตราการตกแจงนับของสำมะโน
2. เพื่อจะได้ทราบสาเหตุและแหล่งที่เกิดความคลาดเคลื่อน

### วิธีการวัดความคลาดเคลื่อนของคุ่มรวม

เลือกท้องที่ตัวอย่าง (Area sample) เพื่อออกไปทำการนับจดทุกครัวเรือนใหม่พร้อมทั้งแจงนับข้อมูลบางรายการของแต่ละคนเพื่อใช้เปรียบเทียบ แล้วนำข้อมูลที่ได้จากการสำรจภายหลังการแจงนับนี้ไปเปรียบเทียบกับข้อมูลที่ได้จากการทำสำมะโนที่ละครัวเรือน และทีละคน ซึ่งความมุ่งหมายของการเปรียบเทียบจะแบ่งครัวเรือนและคนออกเป็น 4 พวก ดังนี้

1. พบทั้งในการสำรจภายหลังการแจงนับและสำมะโน ซึ่งถือว่าเป็นเปรียบเทียบได้
2. พบในการสำรจภายหลังการแจงนับแต่ไม่พบในสำมะโน ซึ่งถือว่าเป็นเปรียบเทียบไม่ได้
3. พบในสำมะโนแต่ไม่พบในการสำรจภายหลังการแจงนับ ซึ่งถือว่าเป็นเปรียบเทียบไม่ได้
4. ไม่พบทั้งในการสำรจภายหลังการแจงนับและสำมะโน ซึ่งไม่ทราบจำนวนที่แน่นอน

แต่อาจหาค่าประมาณได้

นำผลการเปรียบเทียบที่ได้นี้ไปหาค่าความคลาดเคลื่อนของคู่รวม โดยคำนวณหาอัตราความครบถ้วนของการแจงนับ และอัตราการตกแจงนับ โดยถือให้สูงกว่า อัตราว่าการตกแจงนับที่คำนวณได้ไม่ควรเกิน 20%<sup>(4)</sup> แต่ถ้าได้ต่ำกว่า 20% ก็ควรที่จะได้มีการออกไปตรวจสอบงานสนามใหม่ ซึ่งการที่จะประมาณอัตราต่าง ๆ นี้ได้ก็ต้องอาศัยวิธีการประมาณผลจากข้อมูลสองระบบ

#### สูตรที่ใช้ในการวัดความคลาดเคลื่อนของคู่รวม

จากการทำการสำรวจภายหลังการแจงนับ นำข้อมูลในแต่ละเขตตอนหรือหมู่บ้าน มาจัดเป็นรูปตารางได้ดังนี้

PES.	สำมะโน		
	พบ	ไม่พบ	รวม
พบ	a	b	$N_2$
ไม่พบ	c	d	$N - N_2$
รวม	$N_1$	$N - N_1$	N

ข้อมูลจากตารางนำมาคำนวณค่าต่าง ๆ ได้ คือ

#### 1. อัตราความครบถ้วนของการแจงนับของสำมะโน ( w )

$$w = \frac{\text{จำนวนคนหรือครัวเรือนที่เปรียบเทียบได้}}{\text{จำนวนคนหรือครัวเรือนทั้งหมดใน PES}}$$

$$= \frac{a}{N_2} \times 100 \%$$

#### 2. อัตราการตกแจงนับ ( s )

$$s = (100 - w) \%$$

ซึ่งในการวิจัยนี้จะคำนวณอัตราความครบถ้วนของการแจงนับและอัตราการตกแจงนับในระดับเขตตอนและระดับอำเภอ โดยแยกออกเป็นในเขตเทศบาลและนอกเขตเทศบาล

<sup>4</sup>Marks , E.S., Report on mission to Thailand to investigate matching of PES to Census in Thailand. Pilot Census of 1979 ( n.p.: Thailand ), p. 1 .

3. ค่าแปรปรวนของอัตราความครบถ้วนของการแดงนับ  $V(w)$ 

$$V(w) = 100^2 \left( \frac{1}{\hat{N}_2} \right)^2 \left[ V(\hat{a}) + \left( \frac{\hat{a}}{\hat{N}_2} \right)^2 V(\hat{N}_2) - 2 \left( \frac{\hat{a}}{\hat{N}_2} \right) \text{Cov}(\hat{a}, \hat{N}_2) \right] \quad (5)$$

โดยที่

$$\hat{N}_2 = \sum_h N_h \bar{y}_h$$

$$w = \frac{\sum_h N_h \bar{a}_h}{\sum_h N_h \bar{y}_h}$$

$$V(\hat{a}) = \sum_h \sum_g (a_{hg1} - a_{hg2})^2$$

$$V(\hat{N}_2) = \sum_h \sum_g (y_{hg1} - y_{hg2})^2$$

$$\text{Cov}(\hat{a}, \hat{N}_2) = \sum_h \sum_g [(a_{hg1} - a_{hg2})(y_{hg1} - y_{hg2})]$$

ในเมื่อ

$$N_h = \text{จำนวนท้องที่ทั้งหมดในสตรดั้มที่ } h$$

$$\bar{y}_h = \text{ค่าเฉลี่ยของจำนวนคนหรือครัวเรือนที่ถูกนับตลอดท้องที่ใดอย่างในสตรดั้ม } h$$

$$\bar{a}_h = \text{ค่าเฉลี่ยของจำนวนคนหรือครัวเรือนที่เปรียบเทียบได้ต่อท้องที่ใดอย่างในสตรดั้ม } h$$

<sup>5</sup> Bureau of the Census, A Case Study for the 1970 Censuses of Population and Housing; New Florencia (Washington D.C., 1971), p. 78.

$a_{hg_1}$  = จำนวนคนหรือครัวเรือนที่เปรียบเทียบได้ในห้องที่ตัวอย่างที่ 1 ของคู่ที่  $g$   
ในสตราตัม  $h$

$a_{hg_2}$  = จำนวนคนหรือครัวเรือนที่เปรียบเทียบได้ในห้องที่ตัวอย่างที่ 2 ของคู่ที่  $g$   
ในสตราตัม  $h$

$y_{hg_1}$  = จำนวนคนหรือครัวเรือนที่เปรียบเทียบได้ในห้องที่ตัวอย่างที่ 1 ของคู่ที่  $g$   
ในสตราตัม  $h$

$y_{hg_2}$  = จำนวนคนหรือครัวเรือนที่เปรียบเทียบได้ในห้องที่ตัวอย่างที่ 2 ของคู่ที่  $g$   
ในสตราตัม  $h$

#### หมายเหตุ

วิธีในการจับคู่เพื่อคำนวณค่า  $V(a)$  และ  $V(N)$  จะเรียงตามลำดับในบัญชีรายชื่อห้อง  
ที่ตัวอย่าง และจับคู่กันดังนี้ คือ

$$\begin{array}{l} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{array} \quad \begin{array}{l} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{array}$$

ถ้าหากเหลือเศษหนึ่งแห่งจะจับคู่กับรายชื่อห้องที่อยู่เหนือขึ้นมาในบัญชีนั้น

ในการวิจัยนี้จะคำนวณค่าความแปรปรวนของ อัตราความครบถ้วนของการแจกนับเป็น รายอำเภอ โดยแบ่งออกเป็นในเขตเทศบาลและนอกเขตเทศบาล

### การวัดความคลาดเคลื่อนของการแจ้งลักษณะ (Content Evaluation)

วัตถุประสงค์ของการวัดความคลาดเคลื่อนของการแจ้งลักษณะก็เพื่อดูว่าข้อมูลเกี่ยวกับ ลักษณะต่าง ๆ เป็นรายครัวเรือนหรือรายคนที่ได้จากสำมะโนมีความถูกต้องแม่นยำและน่าเชื่อถือได้ มากน้อยแค่ไหน

#### วิธีการวัดความคลาดเคลื่อนของการแจ้งลักษณะ

ใช้ครัวเรือนจากแบบนับจุดของการสำรวจภายหลังการแจกนับในท้องที่ตัวอย่าง เป็นหน่วย แจกนับที่จะต้องออกไปทำการแจกนับซ้ำ โดยจะเลือกตามเฉพาะลักษณะ ที่สำคัญ ๆ จากแบบแจกนับ ของสำมะโนเป็นรายครัวเรือนและรายคน เพื่อนำข้อมูลที่ได้ไปเปรียบเทียบกับข้อมูลของครัวเรือน หรือของบุคคลที่เปรียบเทียบได้ จากสำมะโนลักษณะต่อลักษณะ ซึ่งผลจากการเปรียบเทียบนี้จะได้นำไปหาค่าความคลาดเคลื่อนของการแจ้งลักษณะ โดยจะใช้ค่าความคลาดเคลื่อนสุทธิ (Net error) และค่าความคลาดเคลื่อนทั้งหมด (Gross error) เป็นตัววัด ดังนี้คือ

1. ใช้ความคลาดเคลื่อนสุทธิเป็นตัววัด โดยจะคำนวณออกมาในรูปของอัตราต่าง ๆ คือ
  - ก. อัตราความแตกต่างสุทธิ (Net difference rate)
  - ข. อัตราการเคลื่อนสุทธิ (Net shift rate)

ทั้งสองค่านี้จะใช้วัดดูว่าจำนวนที่ได้จากสำมะโนในแต่ละพวกของแต่ละลักษณะนั้นมาก หรือน้อยกว่าจากการทำสำรวจภายหลังการแจกนับ คือ ถ้าค่าที่คำนวณได้มีค่าเป็นบวก (+) ก็แสดงว่าจำนวนที่ได้จากสำมะโนมากกว่าจากสำรวจภายหลังการแจกนับ โดยคิดเป็นเปอร์เซ็นต์ของจำนวนที่เปรียบเทียบได้ทั้งหมด แต่ถ้ามีค่าเป็นลบ (-) ก็แสดงว่าจำนวนที่ได้จากสำมะโนน้อยกว่า จากการสำรวจภายหลังการแจกนับ แต่อัตราความแตกต่างสุทธิเป็นอัตราที่ใช้กันมากที่สุด

2. ใช้ความคลาดเคลื่อนทั้งหมดเป็นตัววัด โดยจะคำนวณออกมาในรูปของดัชนีต่าง ๆ คือ

ก. ดัชนีของความไม่แนบเนียน (Index of inconsistency) ใช้แสดงสำหรับแต่ละพวกในการกระจาย เป็นค่าดัชนีที่ใช้มากที่สุด

ข. ดัชนีความไม่แนบเนียนแบบแอล-โฟลด์ (L-fold index of inconsistency) ใช้สำหรับการกระจายที่มีมากกว่า 2 พวก ขึ้นไป ซึ่งเป็นค่าเฉลี่ยของดัชนีของแต่ละลักษณะนั่นเอง ทั้งสองค่านี้จะเป็นค่าที่บ่งชี้ความถูกต้องแม่นยำ และความเชื่อถือได้ของวิธีการวัดที่ใช้ หรือใช้ประเมินวิธีการรวบรวมข้อมูลว่าดีพอที่จะทำให้การวัดลักษณะนั้น ๆ สัมเหตุสัมพันธ์หรือไม่ นอกจากนี้ยังมีหน่วยอื่น ๆ ที่ใช้วัดอีก แต่ที่ว่า ไม่ค่อยนิยมใช้กันแพร่หลาย ได้แก่ อัตราการแตกต่างกันรวม (Gross difference rate) อัตราการเคลื่อนรวม (Gross shift rate) และอัตราร้อยละรายงานตรงกัน (Percent identically reported)

กฎเกณฑ์ที่จะใช้ตัดสินว่าค่าดัชนีขนาดไหนจึงจะถือว่าดีค่ามากหรือน้อย อาจกำหนดได้ดังนี้คือ (6)

1. ถ้าค่าดัชนีค่าน้อยกว่า 20% ถือว่ามีค่าเล็ก
2. ถ้าค่าดัชนีค่า 20% - 50% ถือว่ามีค่าปานกลาง
3. ถ้าค่าดัชนีค่ามากกว่า 50% ถือว่ามีค่าใหญ่

ในกรณีที่ค่าดัชนีสำหรับพวกใด หรือลักษณะใด มีค่าใหญ่ แสดงว่า ค่าตอบที่ได้รับในเรื่องนั้น ๆ ไม่น่าเชื่อถือ (Unreliable) ซึ่งควรที่จะมีการปรับปรุงวิธีในการเก็บรวบรวมข้อมูลในเรื่องต่าง ๆ เหล่านี้

---

<sup>6</sup> Bureau of the Census, Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS. Census Match. (A United States Department of Commerce Publication, 1975.), p. 4



สูตรที่ใช้ในการวัดความคลาดเคลื่อนของการแจ้งลักษณะ

จากข้อมูลของการทำสำรวจภายหลังการแจ้งนับ นำมาสร้างตารางลักษณะละ 1 ตาราง

PES. (i = 1, ..., L)	ลักษณะใน (j = 1, ..., L)						รวม
	พวก ที่ 1	พวก ที่ 2	...	พวก ที่ j	...	พวก ที่ L	
พวกที่ 1	$X_{11}$	$X_{12}$	...	$X_{1j}$	...	$X_{1L}$	$X_{1.}$
พวกที่ 2	$X_{21}$	$X_{22}$	...	$X_{2j}$	...	$X_{2L}$	$X_{2.}$
.	.	.	...	.	...	.	.
.	.	.	...	.	...	.	.
พวกที่ i	$X_{i1}$	$X_{i2}$	...	$X_{ij}$	...	$X_{iL}$	$X_{i.}$
.	.	.	...	.	...	.	.
.	.	.	...	.	...	.	.
พวกที่ L	$X_{L1}$	$X_{L2}$	...	$X_{Lj}$	...	$X_{LL}$	$X_{L.}$
รวม	$X_{.1}$	$X_{.2}$	...	$X_{.j}$	...	$X_{.L}$	$n_{..}$

1. อัตราความแตกต่างสำหรับของพวกที่ i :

$$\bar{e}_i = \frac{X_{.j} - X_{i.}}{n_{..}} \times 100 \quad , (i = 1, \dots, L) \quad , (j = 1, \dots, L)$$

, i = j

2. อัตราการเคลื่อนสำหรับของพวกที่ i :

$$\frac{\bar{e}_i}{P_i} = \frac{\left( \frac{X_{.j} - X_{i.}}{n_{..}} \right)}{\left( \frac{X_{i.}}{n_{..}} \right)} \times 100 \quad , (i = 1, \dots, L) \quad , (j = 1, \dots, L)$$

, i = j

3. ดัชนีของความไม่แน่นอนของ พหุคูณที่ 1

$$\hat{I} = \frac{(x_{.j} + x_{i.} - 2 x_{ij})}{\frac{1}{n..} [x_{.j} (n.. - x_{i.}) + x_{i.} (n.. - x_{.j})]} \times 100, \quad (i = 1, \dots, L)$$

$$, \quad (j = 1, \dots, L)$$

$$, \quad i = j$$

4. ดัชนีของความไม่แน่นอนแบบแอค-โพลด์ ของแต่ละลักษณะ

$$\hat{I}_L = \frac{(n.. - \sum_i^L x_{i.j})}{(n.. - \frac{1}{n..} \sum_i^L x_{.j} x_{i.})} \times 100, \quad (i = 1, \dots, L)$$

$$, \quad (j = 1, \dots, L)$$

$$, \quad i = j$$

การคำนวณ 95% ช่วงจำกัดความเชื่อมั่น (Confidence Intervals) ของค่าต่าง ๆ

95% ช่วงจำกัดความเชื่อมั่น ของอัตราความแตกต่างสุทธของพหุคูณที่ 1  $(i = 1, \dots, L)$

$(j = 1, \dots, L)$

$i = j$

95% ขีดจำกัดความเชื่อมั่น คือ

$$\frac{(x_{.j} - x_{i.}) \pm 2 \sqrt{x_{.j} + x_{i.} - 2 x_{ij} + 1}}{n..} \times 100$$

ข้อยกเว้น

ก. ถ้า  $(x_{i.} - x_{ij}) = 0$

ให้ขยายค่าสูงของ ขีดจำกัดความเชื่อมั่น โดยการบวกด้วย  $\frac{2}{n..} \times 100$

ข. ถ้า  $(x_{.j} - x_{ij}) = 0$

ให้ขยายค่าต่ำของ ขีดจำกัดความเชื่อมั่น โดยลบด้วย  $\frac{2}{n..} \times 100$

ค. ถ้าเป็นทั้ง 2 กรณี ก. และ ข.

95% ช่วงจำกัดความเชื่อมั่น คือ

$$- \frac{4}{n_{..}} \times 100 \text{ กับ } + \frac{4}{n_{..}} \times 100$$

95% ช่วงจำกัดความเชื่อมั่น ของดัชนีความไม่แน่นอนของ พวกที่  $i$  ( $i = 1, \dots, L$ )

$$\text{ก. ถ้า } \frac{X_{..j} + X_{i.} - 2X_{ij}}{n_{..}} \leq .10 \quad \begin{matrix} (j = 1, \dots, L) \\ i = j \end{matrix}$$

95% ช่วงจำกัดความเชื่อมั่น คือ

$$\frac{(X_{..j} + X_{i.} - 2X_{ij} + 2) \pm 2\sqrt{X_{..j} + X_{i.} - 2X_{ij} + 1}}{n_{..}} \times 100$$

$$X_{..j} \left( 1 - \frac{X_{i.}}{n_{..}} \right) + X_{i.} \left( 1 - \frac{X_{..j}}{n_{..}} \right)$$

$$\text{ข. ถ้า } \frac{X_{..j} + X_{i.} - 2X_{ij}}{n_{..}} > .10$$

95% ช่วงจำกัดความเชื่อมั่น คือ

$$\frac{(X_{..j} + X_{i.} - 2X_{ij} + 2) \pm 2\sqrt{\frac{1}{n_{..}}(X_{..j} + X_{i.} - 2X_{ij})(n_{..} - X_{..j} + 2X_{ij})}}{n_{..}} \times 100$$

$$X_{..j} \left( 1 - \frac{X_{i.}}{n_{..}} \right) + X_{i.} \left( 1 - \frac{X_{..j}}{n_{..}} \right)$$

95% ช่วงจำกัดความเชื่อมั่น ของดัชนีความไม่แน่นอน แบบ แอล-โพลด์ ของแต่ละลักษณะ

$$\text{ก. ถ้า } \frac{\left( n_{..} - \sum_j X_{ij} \right)}{n_{..}} \leq .10$$

95% ช่วงจำกัดความเชื่อมั่น คือ

$$\frac{(n_{..} - \sum_1^L X_{ij} + 2) \pm 2 \sqrt{n_{..} - \sum_1^L X_{ij} + 1}}{n_{..}}$$

$$\left( n_{..} - \frac{1}{n_{..}} \sum_1^L X_{.j} X_{i.} \right)$$

ข. ถ้า  $\frac{(n_{..} - \sum_1^L X_{ij})}{n_{..}} > .10$

95% ขีดจำกัดความเชื่อมั่น คือ

$$\frac{(n_{..} - \sum_1^L X_{ij} + 2) \pm 2 \sqrt{\frac{1}{n_{..}} (n_{..} - \sum_1^L X_{ij}) \left( \sum_1^L X_{ij} \right)}}{n_{..}}$$

$$\left( n_{..} - \frac{1}{n_{..}} \sum_1^L X_{.j} X_{i.} \right)$$

ค่า 95% ขีดจำกัดความเชื่อมั่น ซึ่งคำนวณได้เหล่านี้จะเป็นค่าที่แสดงว่า Probability ของค่าอัตราความแตกต่างลู่หรือ ดัชนีของความไม่แน่นอน และดัชนีของความไม่แน่นอนแบบ แอล-โพลต์ ที่มีค่าอยู่ระหว่างขีดจำกัดล่าง และขีดจำกัดบน ตามที่คำนวณได้นี้จะเท่ากับ 0.95 หรือจากการทำสำรวจซ้ำ ๆ กัน (Repeated survey) 95% ของค่าวัดเหล่านี้จะอยู่ระหว่างขีดจำกัดบนและขีดจำกัดล่าง