เทคนิคการเลือกตัวอย่าง

โครงการสัมมนาเชิงปฏิบัติการ การเตรียมความพร้อมคณะอนุกรรมการสถิติรายสาขา
ของสำนักงานสถิติแห่งชาติ
ครั้งที่ ๒ วันที่ ๒ มีนาคม ๒๕๕๕ ณ ห้องประชุม ๔๐๑ สำนักงานสถิติแห่งชาติ

หทัยชนก พรรคเจริญ
สำนักนโยบายและวิชาการสถิติ สำนักงานสถิติแห่งชาติ

หัวข้อการบรรยาย

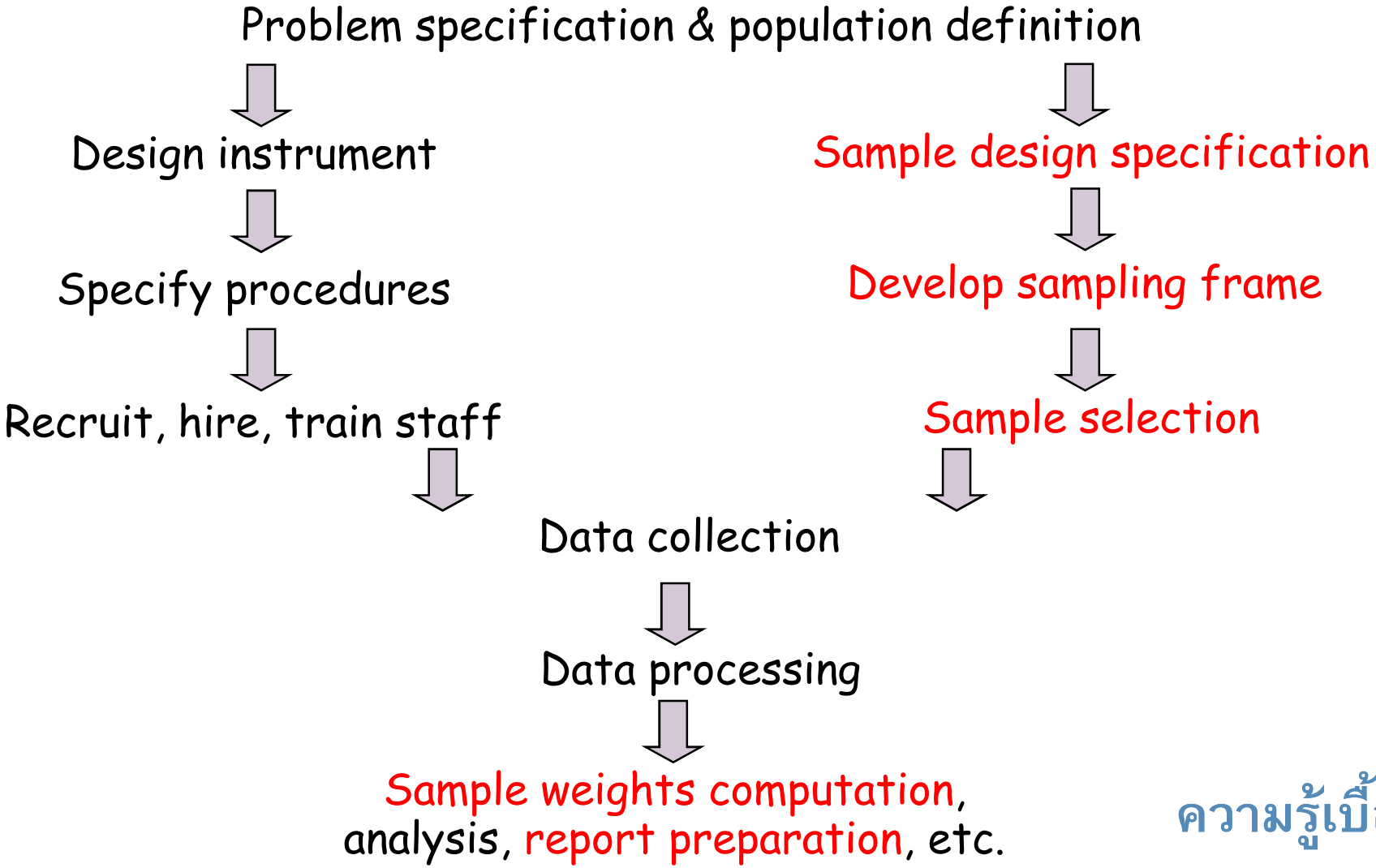
NSO
National Statistical Office
Thailand

1. ความรู้เบื้องต้น
2. วิธีการเลือกหน่วยตัวอย่าง
   - 2.1 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็น
   - 2.2 การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น
3. การกำหนดขนาดตัวอย่าง
4. แผนการสุ่มตัวอย่าง
   - 4.1 แผนการสุ่มตัวอย่างขั้นเดียว
   - 4.2 แผนการสุ่มตัวอย่างหลายขั้น
5. การคำนวณค่าถ่วงน้ำหนัก

NSO
National Statistical Office
Thailand

1. ความรู้เบื้องต้น

Sample Surveys



ความรู้เบื้องต้น

NSO
National Statistical Office
Thailand

ทฤษฎีของการสำรวจด้วยตัวอย่าง

Sample Design

- ควรใช้วิธีการเลือกตัวอย่างอย่างไร
- ควรใช้ขนาดตัวอย่างเท่าไร

Estimation Procedure

- ควรใช้ข้อมูลอะไรประมาณค่า
- การคำนวณค่าถ่วงน้ำหนัก

ความคลาดเคลื่อนของการใช้ตัวอย่างเป็นเท่าไร

ความรู้เบื้องต้น

ความคลาดเคลื่อน

ความคลาดเคลื่อน

(total survey error)

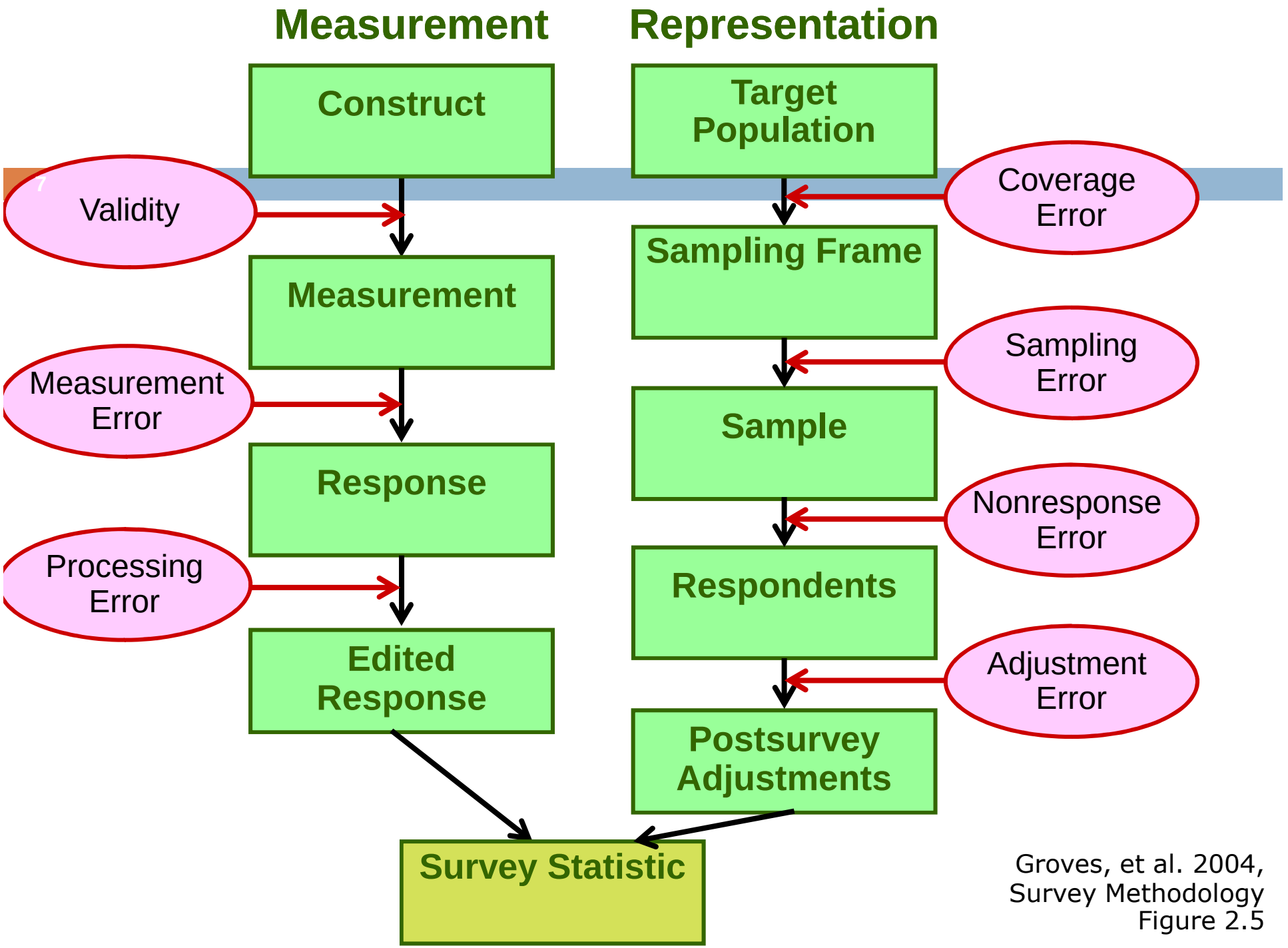
ความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง

(sampling errors)

ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง

(nonsampling errors)

coverage errors, response errors, non-response errors, processing errors, analysis error ...



Groves, et al. 2004,
Survey Methodology
Figure 2.5

การสุ่มตัวอย่าง (sampling?)

- การสุ่มตัวอย่าง คือการเลือกหน่วยตัวอย่างบางหน่วยจากหน่วยตัวอย่างทั้งหมดในประชากร เพื่อเป็นตัวแทนของประชากรที่สนใจศึกษา

Why sampling?

NSO
National Statistical Office
Thailand

- ประหยัดงบประมาณ (reduced cost (economy))
- ประหยัดเวลา (greater speed and timeliness)
- มีคุณภาพ และแม่นยำกว่า (greater quality and accuracy)
- สามารถเก็บรวบรวมข้อมูลรายละเอียดเฉพาะเรื่องได้ (detailed/specialized information)

แบบตัวอย่าง (sample design)

หลักเกณฑ์การเลือกหน่วยตัวอย่าง โดยไม่ให้เกิดความเอนเอียง

- กรอบตัวอย่างต้องเป็นกรอบที่สมบูรณ์
- เลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็น
- ไม่ควรให้พนักงานสนามเลือกหน่วยตัวอย่างเอง
- ไม่ให้มีการเปลี่ยนหน่วยตัวอย่าง
- ปฏิบัติตามวิธีการที่กำหนด

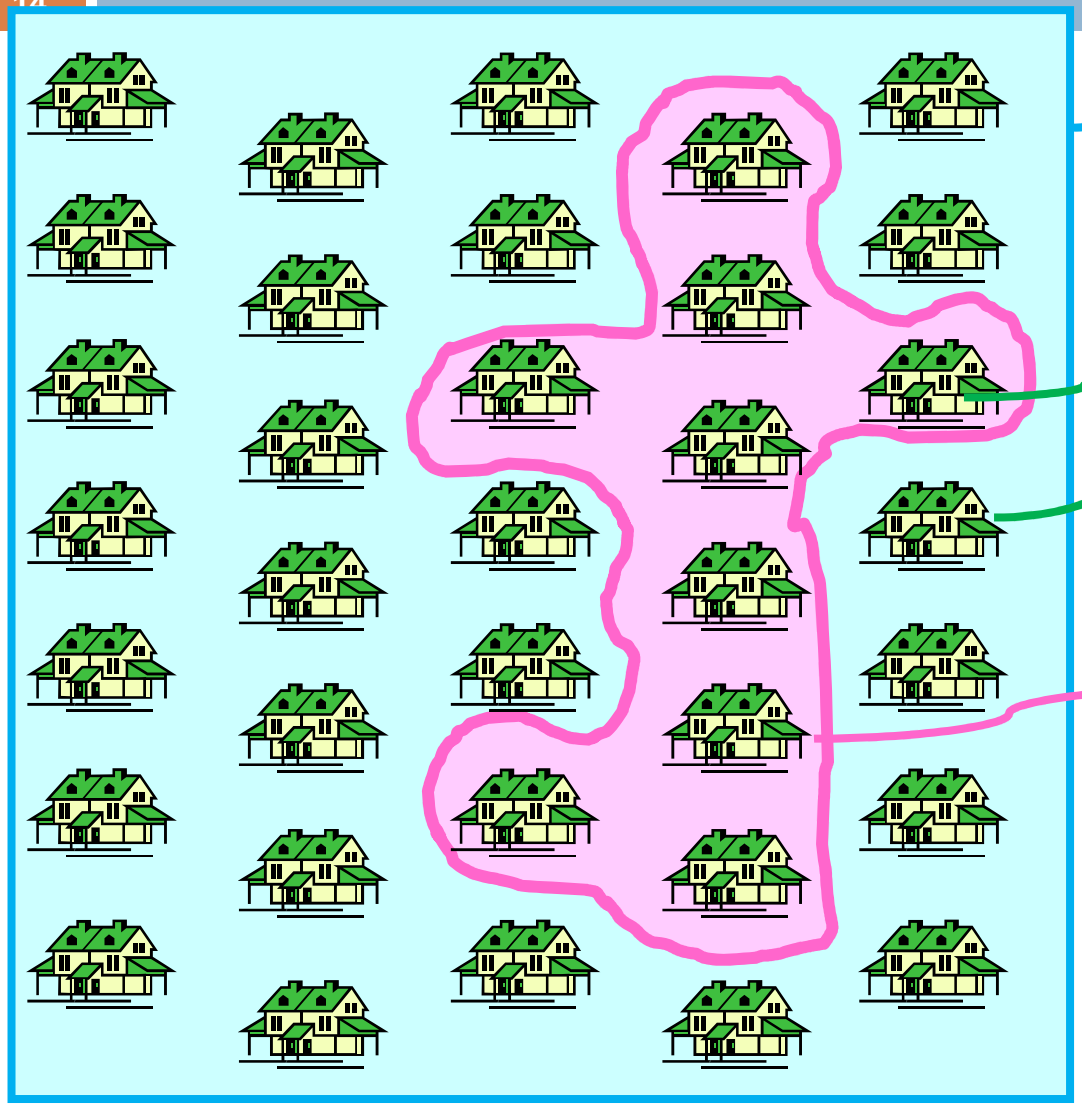
ลักษณะการเลือกหน่วยตัวอย่าง ที่ทำให้เกิดความเอนเอียง

- กรอบตัวอย่างต้องเป็นกรอบที่ไม่สมบูรณ์
- เลือกหน่วยตัวอย่างโดยคิดเอง และไม่ใช้ความน่าจะเป็น
- เลือกหน่วยตัวอย่างที่สนใจเป็นพิเศษให้รวมอยู่ในตัวอย่าง
- เลือกหน่วยตัวอย่างทดแทน
- ไม่เก็บข้อมูลจากหน่วยตัวอย่างที่ถูกเลือกไว้

คำนิยาม

คุ้มรวม (coverage)	พารามิเตอร์ (parameter)
ประชากร (population)	สถิติ (statistic)
หน่วยตัวอย่าง (sampling unit)	ตัวประมาณค่า (estimator)
ตัวอย่าง (sample)	ค่าประมาณ (estimate)
ขนาดตัวอย่าง (sample size)	กรอบตัวอย่าง (sampling frame)

คำนิยาม



ประชากร
(population)

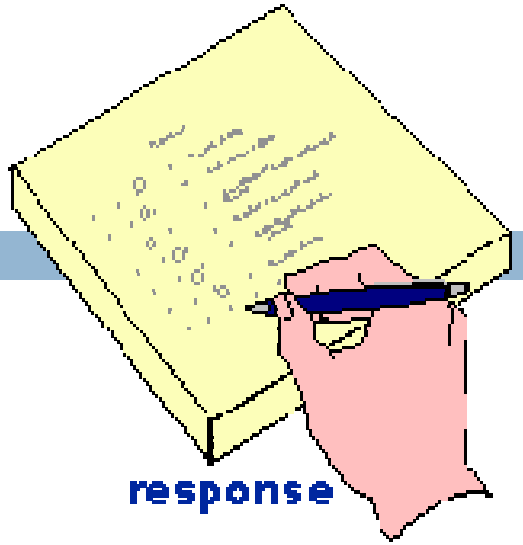
หน่วยตัวอย่าง
(sampling unit)

ตัวอย่าง
(sample)

คำนิยาม

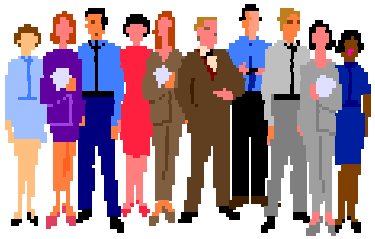
NSO
National Statistical Office
Thailand

Variable



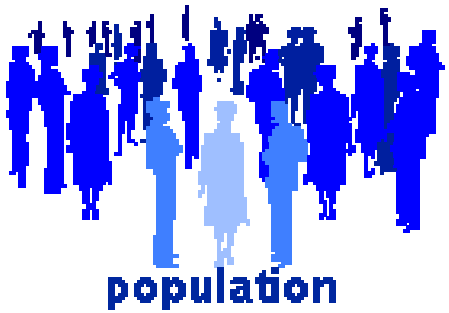
1 2 3 4 5

Statistic



Average = 3.75

Parameter



Average = 3.72

http://www.socialresearchmethods.net/kb/sampstat.php

NSO
National Statistical Office
Thailand

2. วิธีการเลือกหน่วยตัวอย่าง

ลักษณะที่ดีของการออกแบบการเลือกตัวอย่าง

1. ก่อนจะทำการออกแบบการเลือกตัวอย่างจะต้องระบุ **population** ให้ชัดเจน
2. จะต้องจัดหาหรือสร้าง **frame** ที่ดีและทันสมัย
3. โอกาสที่แต่ละ **unit** จะถูกเลือกต้องสามารถคำนวณเป็นตัวเลขได้
4. ควรเป็นแผนแบบที่ง่ายและสามารถปฏิบัติได้ตามที่วางแผนไว้
5. ต้องคำนึงถึงวิสัยสามารถที่จะปฏิบัติได้ ได้แก่ เวลา, ค่าใช้จ่าย, กำลังคน รวมทั้งการประมวลผล
6. จะต้องทำในลักษณะที่ไม่มีความเอนเอียงใด ๆ รวมทั้งในขั้นการสำรวจจริง
7. ควรเลือกแผนแบบที่จะให้ข้อมูลได้มากที่สุด โดยใช้ทรัพยากรที่เท่า ๆ กัน
8. สามารถคำนวณค่าสถิติต่าง ๆ ที่วัดความคลาดเคลื่อนต่าง ๆ ได้
9. จะต้องสามารถใช้หลักของการควบคุมคุณภาพควบคุมขั้นของงานต่าง ๆ ได้

การเลือกหน่วยตัวอย่าง

1. การเลือกหน่วยตัวอย่างโดย**ใช้**ความน่าจะเป็น (probability sampling

- 1.1 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นเท่ากัน (equal probability sampling)
- 1.2 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน (unequal probability sampling)

2. การเลือกหน่วยตัวอย่างโดย**ไม่ใช้**ความน่าจะเป็น (nonprobability sampling

วิธีการเลือกหน่วยตัวอย่าง

การเลือกหน่วยตัวอย่าง

1. การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็น

หมายถึง การเลือกหน่วยตัวอย่างจากหน่วยทุกหน่วยในประชากร ด้วยเทคนิคการสุ่มตัวอย่างตามขนาดตัวอย่างที่กำหนดไว้ โดยหน่วยตัวอย่างแต่ละหน่วยในประชากรสามารถ***คำนวณ***หาโอกาสหรือความน่าจะเป็นที่จะถูกเลือกมาเป็นตัวแทนในตัวอย่างได้

วิธีการเลือก
หน่วยตัวอย่าง

1.1 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นเท่ากัน

การเลือกหน่วยตัวอย่างจำนวน n หน่วยจากทั้งสิ้น N หน่วยในประชากร โดยโอกาสหรือความน่าจะเป็นของแต่ละหน่วยที่ถูกเลือกขึ้นมาเป็นตัวอย่าง (Probability of selection) สามารถคำนวณค่าได้และมีค่าเท่ากันทุกหน่วย

วิธีการเลือกหน่วยตัวอย่างอย่างง่าย (simple random sampling : SRS)

- วิธีการจับฉลาก
- วิธีการใช้ตารางเลขสุ่ม
- วิธีใช้คอมพิวเตอร์ในการสุ่ม

- Sampling with replacement
- Sampling without replacement

วิธีการเลือกหน่วยตัวอย่างแบบมีระบบ (systematic sampling : SYS)

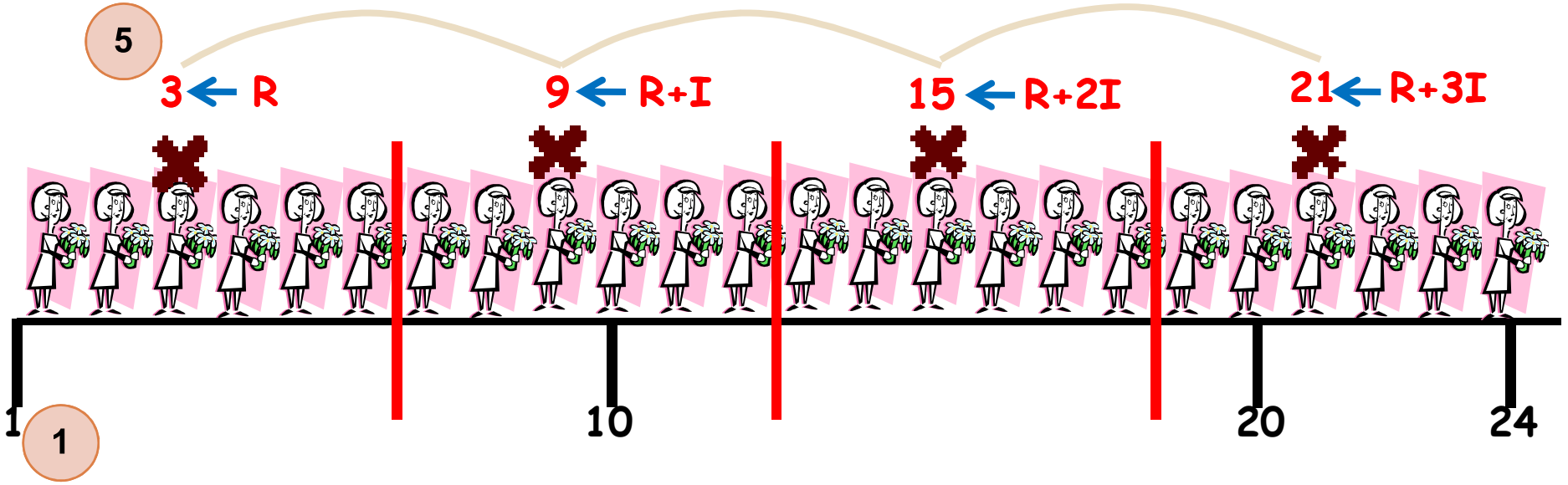
- วิธีการเลือกหน่วยตัวอย่างแบบมีระบบแบบเส้นตรง (linear systematic sampling)
- วิธีการเลือกหน่วยตัวอย่างแบบมีระบบแบบวงกลม (circular systematic sampling)

วิธีการเลือกหน่วยตัวอย่าง

วิธีการเลือกหน่วยตัวอย่างแบบมีระบบ (systematic sampling : SYS)

วิธีการเลือกหน่วยตัวอย่างแบบมีระบบแบบเส้นตรง (linear systematic sampling)

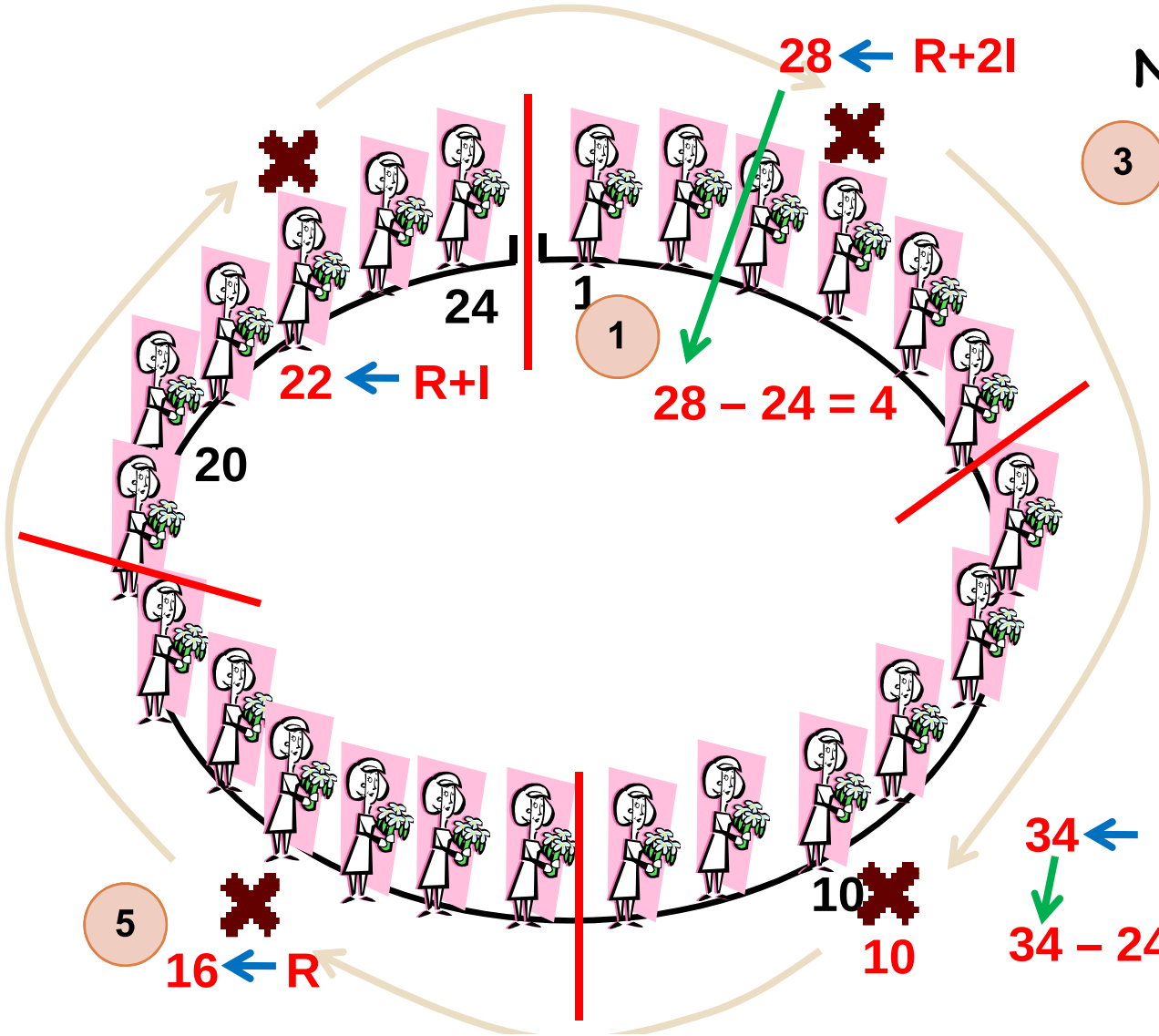
N = 24 (2) n = 4 (3) I = N/n = 24/4 = 6 (4) 0<R≤4 => 3



Sampling Fraction = n/N = 4/24 ⇨ 1 : 6

วิธีการเลือกหน่วยตัวอย่างแบบมีระบบ (systematic sampling : SYS)

วิธีการเลือกหน่วยตัวอย่างแบบมีระบบแบบวงกลม (cercular systematic sampling)



2 N = 24 n = 4

3 I = N/n = 24/4 = 6

4 0<R≤24 => 16

1.1 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นเท่ากัน

วิธีการเลือกหน่วยตัวอย่างอย่างง่าย (simple random sampling : SRS)

ข้อดี	ข้อเสีย
1. ง่าย สะดวกต่อการนำไปใช้	1. ถ้ากรอบตัวอย่างมีข้อมูลสนับสนุนอื่น ๆ การใช้การเลือกตัวอย่างอย่างง่ายจะมีประสิทธิภาพน้อยกว่าการเลือกตัวอย่างด้วยวิธีอื่น
2. ไม่ต้องการข้อมูลสนับสนุน (auxiliary) ในกรอบตัวอย่าง ต้องการเพียงข้อมูลครบทุกหน่วย และที่อยู่	2. ค่าใช้จ่ายสูงในกรณีที่ใช้พนักงานไปสัมภาษณ์
3. ไม่จำเป็นต้องมีการพัฒนาในเชิงเทคนิคที่ซับซ้อน	3. เป็นไปได้ที่จะได้ตัวแทนที่ไม่ดี ถ้าชุดตัวอย่างที่เลือกได้ไม่กระจาย

1.1 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นเท่ากัน

วิธีการเลือกหน่วยตัวอย่างแบบมีระบบ (systematic sampling : SYS)

ข้อดี	ข้อเสีย
1. เป็นตัวแทนของการเลือกตัวอย่างอย่างง่ายในกรณีที่ไม่มีกรอบตัวอย่าง	1. ถ้ากรอบตัวอย่างมีข้อมูลสนับสนุนอื่น ๆ การใช้การเลือกตัวอย่างอย่างง่ายจะมีประสิทธิภาพน้อยกว่าการเลือกตัวอย่างด้วยวิธีอื่น
2. ไม่ต้องการข้อมูลสนับสนุน (auxiliary) ในกรอบตัวอย่าง	2. ในกรณีที่ไม่มีกรอบตัวอย่าง จะทำให้ไม่ทราบ n จนกว่าจะเก็บข้อมูลเสร็จ
3. ตัวอย่างที่เลือกได้มีการกระจายมากกว่า SRS	3. ถ้า n ไม่สามารถหาร N ได้ลงตัว ทำให้ได้ n ที่เท่ากับที่กำหนดไว้ โดยอาจเลี่ยงไปใช้การเลือกแบบมีระบบแบบวงกลม
4. การคำนวณค่าประมาณง่าย ไม่ซับซ้อน	
5. ง่ายและสะดวกกว่า SRS	

1.2 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน

หมายถึง การเลือกหน่วยตัวอย่างจำนวน *n* หน่วยจากทั้งสิ้น *N* หน่วยในประชากร โดยโอกาสหรือความน่าจะเป็นของแต่ละหน่วยที่ถูกเลือกขึ้นมาเป็นตัวแทนในตัวอย่างสามารถคำนวณค่าได้ และมีค่าไม่เท่ากัน ขึ้นอยู่กับหน่วยวัดขนาด (measure of size : MOS) ของหน่วยนั้น ๆ เรียกว่า probability proportional to size (PPS)

วิธีการเลือกหน่วยตัวอย่างแบบ PPS - random

วิธีการเลือกหน่วยตัวอย่างแบบ PPS - systematic

วิธีการเลือกหน่วยตัวอย่าง

1.2 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน

NSO National Statistical Office Thailand

วิธีการเลือกหน่วยตัวอย่างแบบ PPS - random

2 N = 16 n = 4

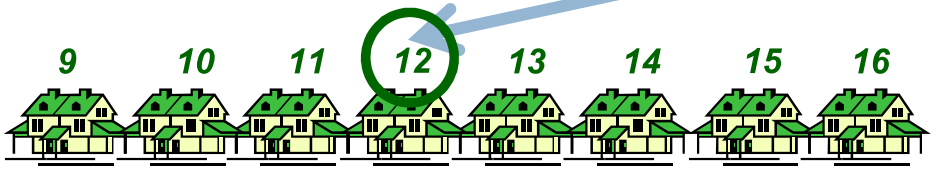
7 เตรียมฉลาก 1 – 60 และเลือกเลขสุ่มจากการจับฉลากหรือตารางเลขสุ่ม จำนวน n ตัว

1

3 จำนวนสมาชิก M_i = MoS

3	4	2	3	1	6	5	3
3	7	9	12	13	19	24	28

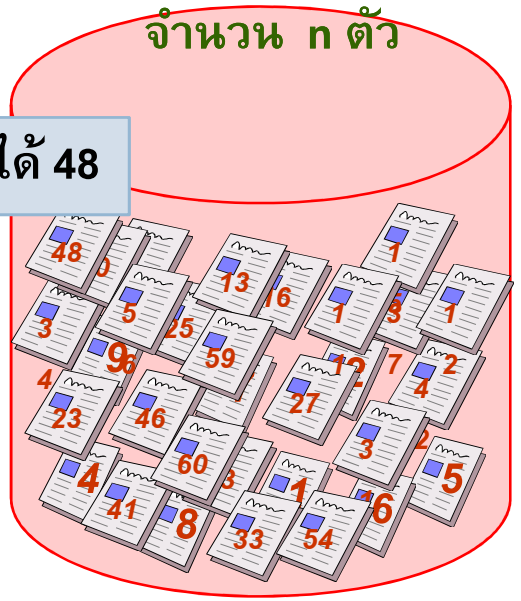
4 ผลบวกสะสม



จำนวนสมาชิก

3	7	2	9	1	2	5	3
31	38	40	49	50	52	57	60

สุ่มได้ 48



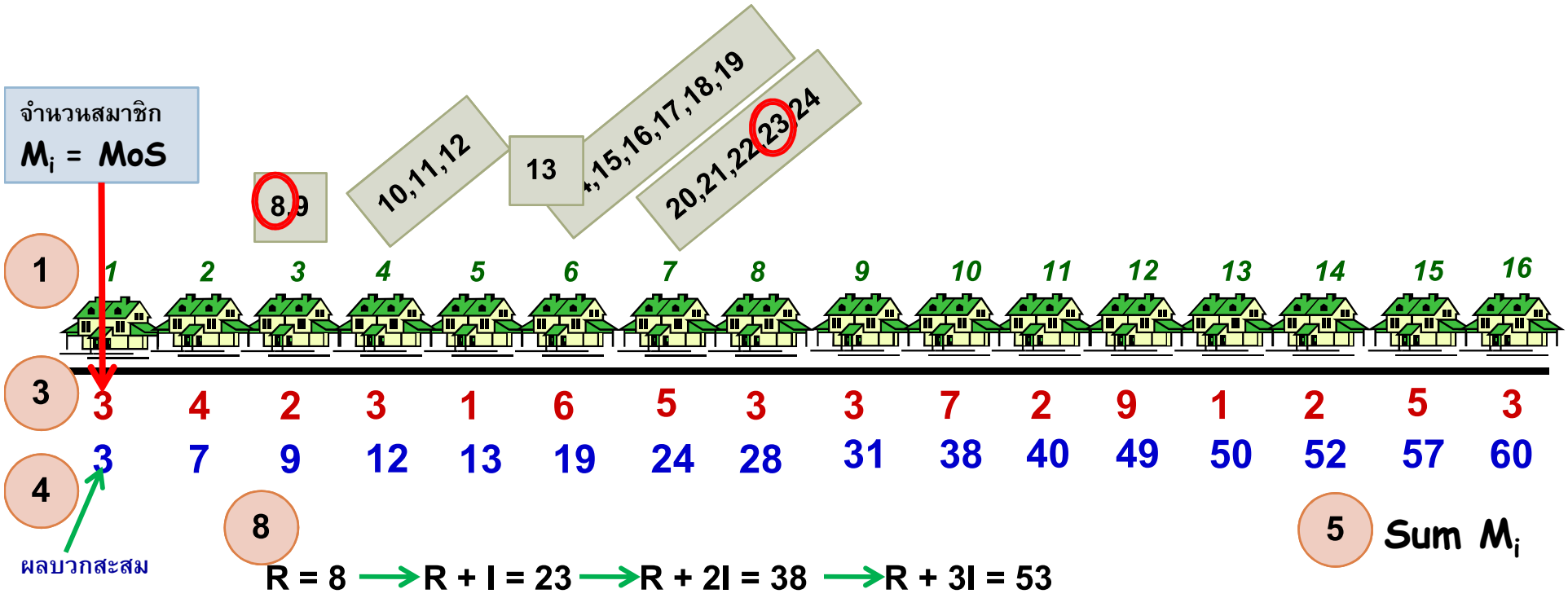
6 Sum M_i

1.2 การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน

NSO
National Statistical Office
Thailand

วิธีการเลือกหน่วยตัวอย่างแบบ PPS - systematic

N = 16 (2) n = 4 (6) I = sum(M_i)/n = 60/4 = 15 (7) 0<R≤15 => 8



2. การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น

- ไม่ทราบจำนวนประชากรที่แท้จริง หรือไม่มีกรอบตัวอย่างที่สมบูรณ์ ทำให้ไม่สามารถใช้การเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็นได้
- หน่วยตัวอย่างมีโอกาสถูกเลือกไม่เท่ากัน บางหน่วยตัวอย่างมีโอกาสถูกเลือกมากกว่าหนึ่งครั้ง หรือบางหน่วยตัวอย่างไม่มีโอกาสที่จะถูกเลือก
- คำนึงถึงความสะดวกทั้งทางด้านเวลา กำลังคน และงบประมาณ รวมทั้งวิธีการเก็บรวบรวมข้อมูลของนักสถิติเป็นหลัก จึงทำให้ไม่ทราบความน่าจะเป็นที่หน่วยแต่ละหน่วยในประชากรจะถูกเลือกเป็นตัวอย่าง
- ไม่สามารถอ้างอิงหรืออนุมานไปยังประชากรที่ต้องการศึกษาได้

วิธีการเลือกหน่วยตัวอย่าง

2. การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น

- วิธีการเลือกหน่วยตัวอย่างโดยบังเอิญ (accidental sampling) หรือ การเลือกหน่วยตัวอย่างตามสะดวก (convenience sampling)
- วิธีการเลือกหน่วยตัวอย่างแบบโควต้า (quota sampling)
- วิธีการเลือกหน่วยตัวอย่างแบบเจาะจง (purposive sampling) หรือ การเลือกหน่วยตัวอย่างแบบใช้วิจารณญาน (judgment sampling)

2. การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น

- วิธีการเลือกหน่วยตัวอย่างโดยบังเอิญ (accidental sampling) หรือการเลือกหน่วยตัวอย่างตามสะดวก (convenience sampling)

เป็นการเลือกหน่วยตัวอย่างที่ไม่มีหลักเกณฑ์ นั่นคือเลือกใครก็ได้ที่สามารถให้ข้อมูลได้แต่ต้องอยู่ในคุ้มรวมของประชากรที่สนใจศึกษา เช่น พนักงานเก็บรวบรวมข้อมูลยืนอยู่ประตูหน้าห้างสรรพสินค้าเพื่อสัมภาษณ์ผู้คนที่ผ่านไปมาบริเวณนั้น

2. การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น

- วิธีการเลือกหน่วยตัวอย่างแบบโควต้า (quota sampling)

- เป็นการเลือกหน่วยตัวอย่างที่พบบ่อยที่สุดในการเลือกตัวอย่างโดยไม่ใช้ความน่าจะเป็น
- ทำการจำแนกประชากรออกเป็นส่วนย่อย ๆ ก่อน (strata) โดยตัวแปรที่ใช้ในการจำแนกควรจะมีความสัมพันธ์กับตัวแปรที่สนใจศึกษา เช่น เพศ อายุ ระดับการศึกษา หรือรายได้
- จากนั้นพิจารณาขนาดตัวอย่างของแต่ละส่วนย่อย เพื่อกำหนดเป็นโควตา หรือจะเรียกได้ว่าเป็นการเลือกหน่วยตัวอย่างโดยคำนึงถึงสัดส่วนองค์ประกอบของประชากร
- แล้วในแต่ละกลุ่มใช้วิธีการเลือกหน่วยตัวอย่างแบบบังเอิญ นั่นคือเจอใครก็เลือกจนครบตามจำนวนที่ต้องการ

2. การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น

- **วิธีการเลือกหน่วยตัวอย่างแบบเจาะจง (purposive sampling) หรือการเลือกหน่วยตัวอย่างแบบใช้วิจารณญาน (judgment sampling)**

เป็นการเลือกตัวอย่างโดยใช้ดุลพินิจและการตัดสินใจของนักสถิติเป็นหลักในการพิจารณาเลือกตัวอย่าง ว่ามีลักษณะสอดคล้องหรือเป็นตัวแทนที่จะศึกษาได้หรือไม่ เป็นไปตามวัตถุประสงค์ของการสำรวจหรือไม่ ทั้งนี้นักสถิติต้องเป็นผู้ที่มีความรอบรู้ ความชำนาญและประสบการณ์

วิธีการเลือกหน่วยตัวอย่าง

เปรียบเทียบการเลือกหน่วยตัวอย่างโดยใช้และไม่ใช้ความน่าจะเป็น

	ใช้ความน่าจะเป็น	ไม่ใช้ความน่าจะเป็น
วิธีการ	• random • systematic sampling • PPS sampling	• purposive sampling • quota sampling • accidental sampling
ข้อดี	• ไม่ลำเอียง/เอนเอียง • ทุกหน่วยมีโอกาสถูกเลือกเป็นหน่วยตัวอย่าง • อ้างอิงไปยังประชากรได้ • ควบคุมความคลาดเคลื่อนที่เกิดจากการเลือกตัวอย่างได้ (sampling error)	• สะดวก • รวดเร็ว • ประหยัดค่าใช้จ่าย
	• มีขั้นตอนที่ยุ่งยาก	• ลำเอียง/เอนเอียง • ไม่สามารถอ้างอิงไปยังประชากรได้ จะสามารถสรุปอยู่เพียงขอบเขตของกลุ่มตัวอย่างเท่านั้น • หน่วยตัวอย่างที่ได้นั้นขึ้นอยู่กับการตัดสินใจของนักสถิติ และองค์ประกอบบางตัวไม่สามารถควบคุมได้ ดังนั้นไม่มีวิธีการทางสถิติที่จะมาคำนวณความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่างได้ (sampling error)

NSO
National Statistical Office
Thailand

3. การกำหนดขนาดตัวอย่าง

สิ่งที่ต้องพิจารณาในการกำหนดขนาดตัวอย่าง

- ประชากรที่ต้องการศึกษา
- ระดับการนำเสนอผลการสำรวจ
- พารามิเตอร์ที่ต้องการศึกษา (ค่าเฉลี่ย, ยอดรวม, ร้อยละ)
- ความแปรปรวน (variance) ของตัวแปรที่ต้องการศึกษา
- ค่าสัมประสิทธิ์ความผันแปร (Coefficient of variation)
- ระดับความเชื่อมั่น (Confidence Level)
- ขนาดของความคลาดเคลื่อนที่ยอมรับได้ (Permissible Error)

สิ่งที่ต้องพิจารณาหลังจากได้ขนาดตัวอย่าง

- งบประมาณและเวลา
- อื่น ๆ เช่น วิธีการเก็บรวบรวมข้อมูล Response Rate ที่คาดหวัง จำนวนพนักงานเก็บข้อมูลที่มี เป็นต้น

ประชากรที่ต้องการศึกษา

- ต้องกำหนดให้ชัดเจน
- ข้อสังเกต...
  - ขนาดตัวอย่างไม่ได้เพิ่มขึ้นเป็นสัดส่วนกับขนาดของประชากร
- บทบาทของขนาดประชากรที่มีต่อการกำหนดขนาดตัวอย่าง
  - มีบทบาทมากเมื่อประชากรมีขนาดเล็ก
  - มีบทบาทน้อยลงเมื่อประชากรมีขนาดใหญ่

ระดับการนำเสนอผลการสำรวจ

- จำเป็นต้องกำหนดให้แน่ชัดเกี่ยวกับระดับการนำเสนอผลการสำรวจ
  - ระดับประเทศ
  - ระดับภาค
  - ระดับจังหวัด
  - ระดับกลุ่ม
- ขนาดตัวอย่างที่กำหนดจะต้องเพียงพอต่อการนำเสนอผลในแต่ละระดับ

พารามิเตอร์ที่ต้องการศึกษา

- พารามิเตอร์ที่ต้องการศึกษา เช่น
  - ยอดรวมของประชากร
  - ค่าเฉลี่ยของประชากร
  - ค่าสัดส่วนของประชากร
- ข้อสังเกต
  - ถ้ามีพารามิเตอร์ที่ต้องการศึกษาหลายลักษณะ จะต้องนำพารามิเตอร์ทุกตัวมาคำนวณขนาดตัวอย่าง และพิจารณาขนาดตัวอย่างที่มีขนาดใหญ่ที่สุด (ถ้าเป็นไปได้)

ความแปรปรวนของตัวแปรที่ต้องการศึกษา

- ถ้าประชากรที่ต้องการศึกษามีลักษณะที่แตกต่างกันมาก (แปรปรวนมาก) ขนาดตัวอย่างที่ใช้ก็จะใหญ่กว่า ประชากรที่มีความแตกต่างกันน้อย
- ความแปรปรวนของประชากร อาจได้จาก
  - ข้อมูลในอดีตที่เคยมีผู้สำรวจไว้แล้ว
  - ข้อมูลจากสำมะโน
  - ข้อมูลจากการสำรวจล่วงหน้า
  - ข้อมูลจากการสำรวจจริง

ค่าสัมประสิทธิ์ความผันแปร (Coefficient of Variation: CV)

- CV คือ อัตราส่วนระหว่าง Standard deviation ของค่าประมาณ ต่อค่าประมาณ
- ค่าประมาณที่มีค่า CV น้อย อาจสรุปได้ว่า ค่าประมาณนั้นมีความน่าเชื่อถือกว่าค่าประมาณที่มีค่า CV มาก

ระดับความเชื่อมั่น (Confidence Level)

- แสดงถึงโอกาสที่ค่าประมาณที่ต้องการศึกษาจะมีความถูกต้องมากน้อยเพียงใด เช่น
  - ระดับความเชื่อมั่น 95%
  - ระดับความเชื่อมั่น 99%
- ข้อควรจำ !!!
  - ถ้าต้องการระดับความเชื่อมั่นสูงขึ้น ขนาดตัวอย่างต้องเพิ่มขึ้น

ขนาดของความคลาดเคลื่อนที่ยอมรับได้ (Permissible Error)

- เรียกอีกอย่างหนึ่งว่า ช่วงของค่าประมาณ
  - เป็นค่าที่แสดงถึงความแตกต่างของค่าประมาณจากตัวอย่างกับค่าพารามิเตอร์ของประชากร
  - เป็นค่าที่ระบุว่าผู้สำรวจต้องการให้ค่าประมาณแตกต่างไปจากค่าพารามิเตอร์ด้วยช่วงที่ยอมรับได้เท่าไหร่ โดยปกติจะแสดงในรูปบวก/ลบ เช่น +/-3
- ข้อควรจำ !!!
  - ยิ่งช่วงความคลาดเคลื่อนที่ยอมรับได้แคบลง ขนาดตัวอย่างจะมีขนาดใหญ่ขึ้น

งบประมาณและเวลา

- ส่วนใหญ่ขนาดตัวอย่างที่ต้องการมักจะมีขนาดใหญ่เกินกว่างบประมาณที่มี
  - ถ้าไม่สามารถหางบประมาณเพิ่มเติมได้ ก็ต้องยอมลดขนาดตัวอย่างลง
- เช่นเดียวกับเรื่องของเวลา
  - ถ้าไม่สามารถขยายคาบการปฏิบัติงานได้ ก็ต้องยอมลดขนาดตัวอย่างลง

ปัจจัยอื่น ๆ

- วิธีการเก็บรวบรวมข้อมูล
- ระยะเวลาที่ทำการสำรวจ
- **Response Rate** ที่คาดหวัง
- จำนวนพนักงานเก็บข้อมูลที่มี

Expected Response Rate

- Non-response can occur for many reasons
  - Members of population being surveyed may not be available
  - Inability to make contact
  - Unexpected illness
  - Refuse to answer questionnaires
- Non-response varies by the collection data method chosen
  - Mail survey
  - Face-to-face interviewing
- Response rate can be estimated
  - with the help of a pilot survey
  - from past experience with similar surveys

สูตรการคำนวณขนาดตัวอย่าง

ค่าประมาณ $(\hat{\theta})$	ขนาดตัวอย่าง (n)	
	Absolute Error : $\left\|\hat{\theta}-\theta\right\| \le E$	**Relative Error :** $\left\|\frac{\hat{\theta}-\theta}{\theta}\right\| \le E'$
1. ยอดรวม $(\hat{X})$	$n = \frac{N^2k^2\sigma_X^{\,2}}{Nk^2\sigma_X^{\,2}+E^2}$	$n_{rel} = \frac{Nk^2V^2}{k^2V^2+NE'^2}$
2. ค่าเฉลี่ย $(\bar{\hat{X}})$	$n = \frac{Nk^2\sigma_X^{\,2}}{k^2\sigma_X^{\,2}+NE^2}$	$n_{rel} = \frac{Nk^2V^2}{k^2V^2+NE'^2}$
3. สัดส่วน $(\hat{P})$	$n = \frac{Nk^2PQ}{k^2PQ+NE^2}$	$n_{rel} = \frac{Nk^2Q}{k^2Q+NPE'^2}$

สรุปขั้นตอนในการกำหนดขนาดตัวอย่าง

1. คำนวณขนาดตัวอย่างเริ่มต้น
2. ถ้าไม่ได้ใช้ SRS ให้ปรับขนาดตัวอย่างให้สอดคล้องกับแผนการสุ่มตัวอย่างที่ใช้ หรือ ปรับด้วย Design Effect นั่นเอง
3. ปรับขนาดตัวอย่างขั้นสุดท้ายด้วยค่า Response Rate ที่คาดหวัง

NSO
National Statistical Office
Thailand

4. แผนการสุ่มตัวอย่าง

แผนการสุ่มตัวอย่าง

- แผนการสุ่มตัวอย่างขั้นเดียว (single stage sampling)
  - แผนการสุ่มตัวอย่างอย่างง่าย (simple random sampling)
  - แผนการสุ่มตัวอย่างแบบมีระบบ (systematic sampling)
  - แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ (stratified sampling)
  - แผนการสุ่มตัวอย่างแบบแบ่งกลุ่ม (cluster sampling)
- แผนการสุ่มตัวอย่างหลายขั้น (multi stage sampling)

แผนการสุ่มตัวอย่างอย่างง่าย (simple random sampling)

SRS เป็นวิธีการเลือกตัวอย่างที่ง่ายที่สุดคือเป็นการเลือกหน่วยตัวอย่างเพียงขั้นเดียว โดยใช้วิธีการเลือกตัวอย่างแบบ random โดยที่ทุกหน่วยในประชากรมีโอกาสที่จะถูกเลือกเท่ากันหมด

หลักการใช้แผนการสุ่มตัวอย่างแบบ SRS

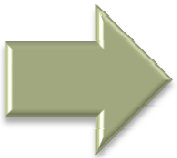
ควรใช้กับประชากร ซึ่งหน่วยต่าง ๆ มีลักษณะคล้ายคลึงกันเป็นส่วนใหญ่ และการกระจายของข้อมูลมีไม่มากนัก

มีหลายวิธีที่ใช้ในการเลือกตัวอย่าง เช่นการจับสลาก การใช้ตารางเลขสุ่ม หรือ random number generators

แผนการสุ่มตัวอย่างอย่างง่าย

(simple random sampling)

ประชากร



ตัวอย่าง

แผนการสุ่มตัวอย่างแบบมีระบบ (systematic sampling)

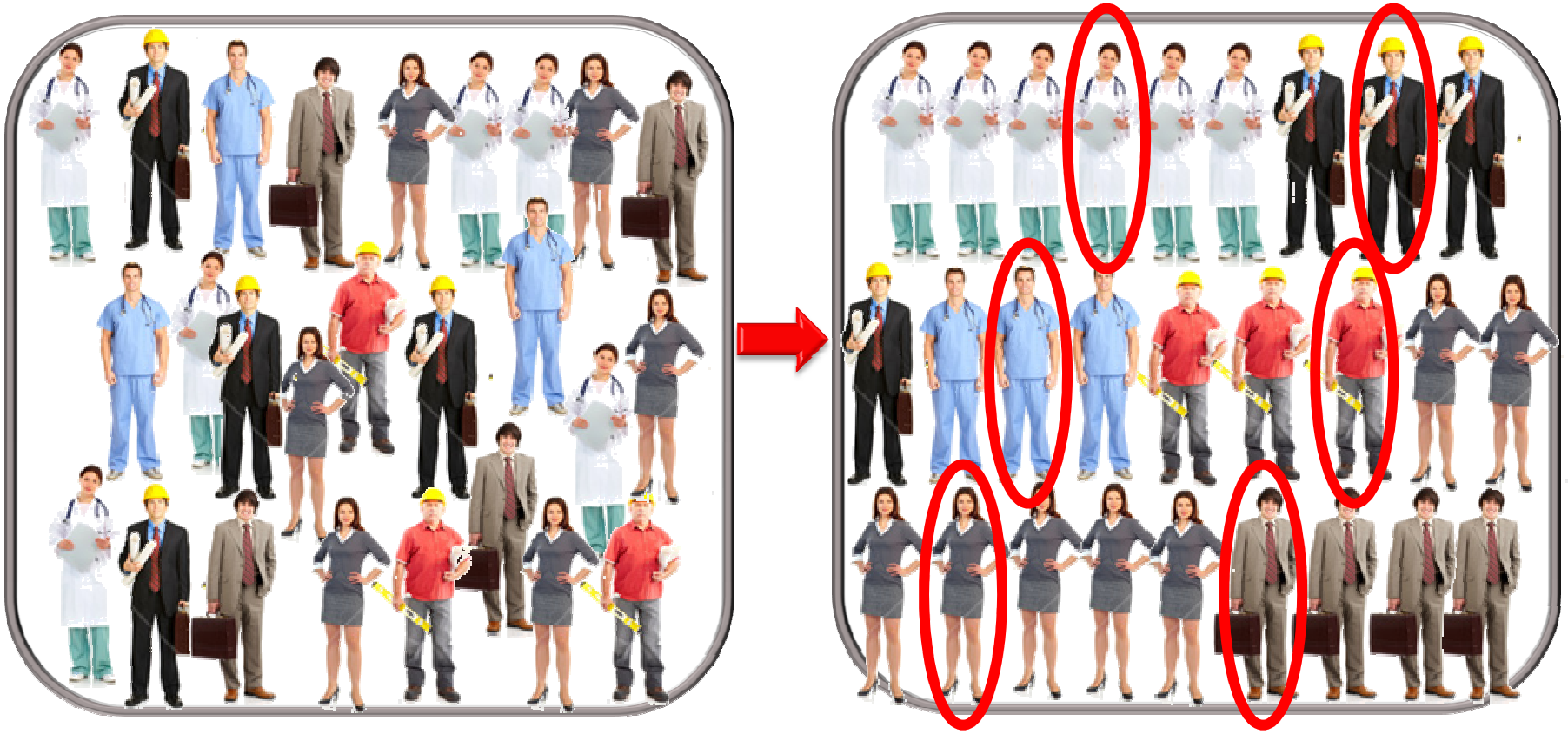
เป็นแผนการการสุ่มตัวอย่างที่คล้ายกับ SRS คือเป็นการเลือกหน่วยตัวอย่างขนาด n หน่วยจากประชากรจำนวน N หน่วยโดยทำการเลือกหน่วยตัวอย่างหน่วยแรกแบบสุ่ม จากหน่วยที่ 1 ถึง หน่วยที่ k และ ต่อจากนั้นจะเลือกหน่วยตัวอย่างต่อไปทุก ๆ k หน่วย จนกระทั่งครบ n หน่วยตามที่ต้องการ

ข้อดี

- ★ เป็นวิธีที่ง่ายในการเลือกหน่วยตัวอย่าง เพราะว่าทำการสุ่มเลขขึ้นมาแค่ 1 ครั้ง ซึ่งเสียเวลาน้อย
- ★ สะดวกในการปฏิบัติงานเมื่อทำการเลือกหน่วยตัวอย่างที่งานสนาม
- ★ การอบรมพนักงานสำรวจจะทำได้สะดวกกว่าแผนแบบการสุ่มตัวอย่างแบบอื่น ๆ
- ★ มีประสิทธิภาพสูงเมื่อประชากรมีการเรียงลำดับของหน่วยตัวอย่าง

แผนการสุ่มตัวอย่างแบบมีระบบ

(systematic sampling)



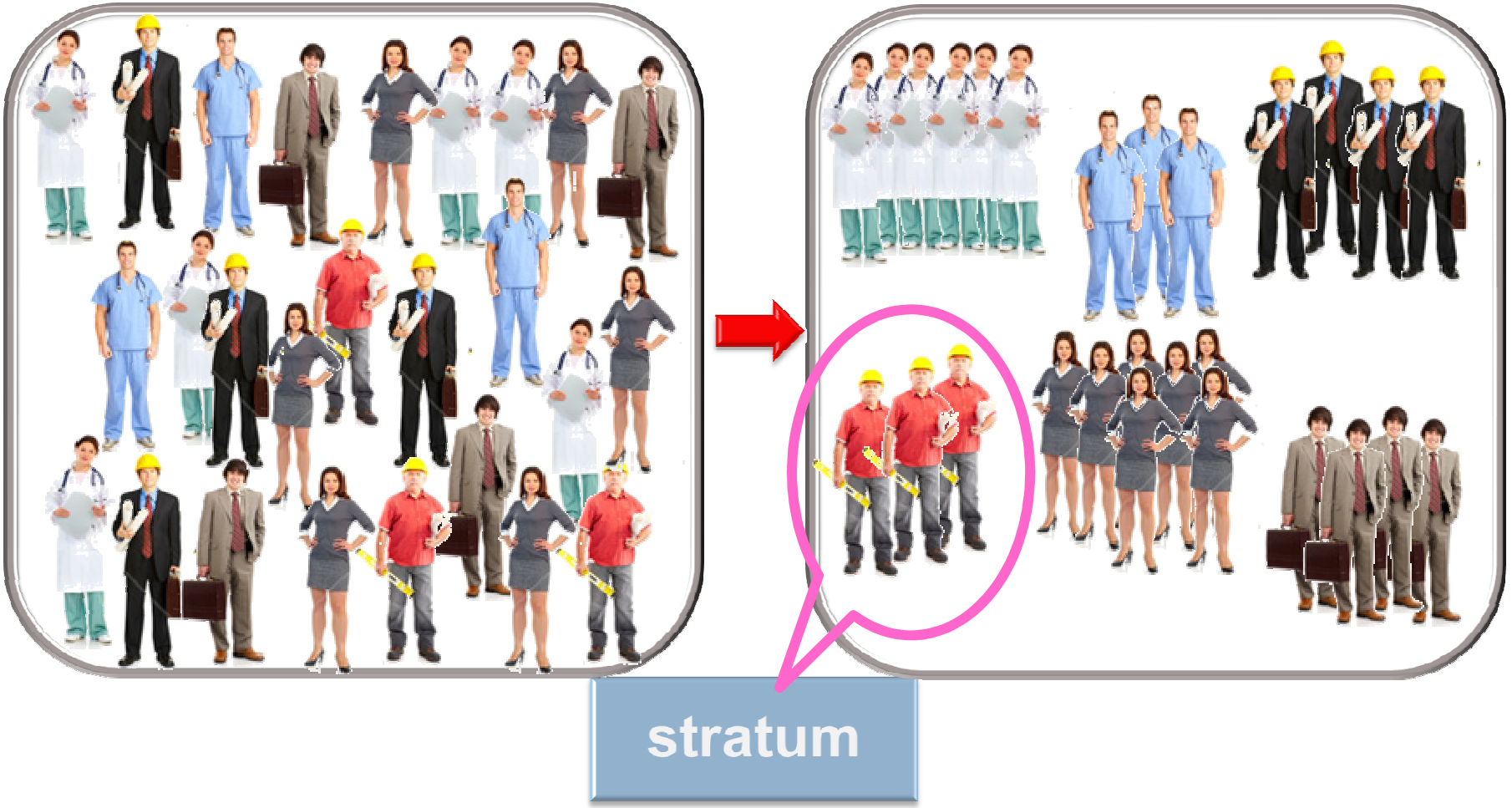
แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ (stratified sampling)

คือ แผนการเลือกตัวอย่างจากประชากรที่มีการแบ่งออกเป็นชั้นภูมิ (stratum) แล้วเลือกตัวแทนของประชากรในแต่ละชั้นภูมิขึ้นมาตามขนาดตัวอย่างที่กำหนด เพื่อเป็นตัวอย่างในการสำรวจ

วิธีการแบ่งประชากรออกเป็นชั้นภูมิ เรียกว่า stratification แต่ละชั้นภูมิของประชากรที่แบ่งออกไปเรียกว่า stratum หลักสำคัญในการแบ่งก็คือ

- ให้หน่วยที่อยู่ในชั้นภูมิเดียวกันควรมีความคล้ายคลึงกัน (homogeneity within stratum) มากที่สุด
- แต่มีความแตกต่างกันระหว่างชั้นภูมิมากที่สุด (heterogeneity between stratum)

แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ (stratified sampling)



แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ (stratified sampling)

- เมื่อต้องการประมาณค่าในแต่ละส่วนย่อย ๆ ของประชากรแยกออกจากกัน
- เมื่อมีกรอบตัวอย่างในแต่ละส่วนย่อย ๆ ของประชากร
- เมื่อในแต่ละส่วนย่อยในประชากรมีการเลือกตัวอย่างที่แตกต่างกัน และง่ายต่อการบริหารจัดการในการเก็บรวบรวมข้อมูล

หลักการใช้แผนแบบการสุ่มตัวอย่างแบบมีชั้นภูมิ จะต้องคำนึงถึง

1. ข้อมูลที่ใช้ในการแบ่งชั้นภูมิ (stratum)
2. ขนาดและขอบเขตของแต่ละชั้นภูมิ
3. จำนวนชั้นภูมิ
4. การจัดสรรขนาดตัวอย่างในแต่ละชั้นภูมิ

แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ (stratified sampling)

Stratification variables: ตัวอย่าง

- Household surveys
  - Area: urban / rural
  - Location: region, province, district
- Establishment surveys
  - ISIC, NACE
  - Employment size
  - Output size
- Multiple stratification variables
  - Industry and size
  - Region and urban/rural

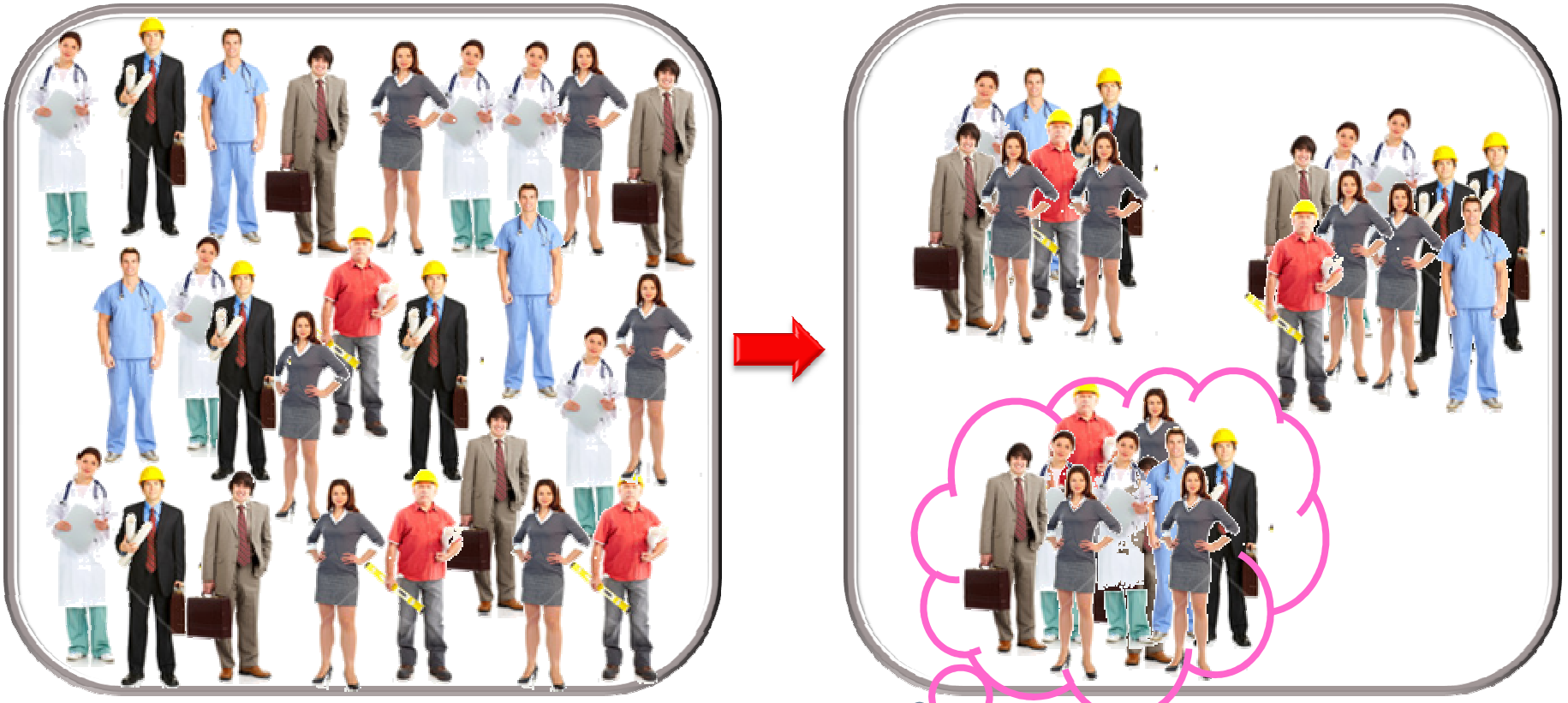
แผนการสุ่มตัวอย่างแบบแบ่งกลุ่ม (cluster sampling)

เป็นวิธีการสุ่มตัวอย่างที่มีการรวมหน่วยตัวอย่างเข้าไว้เป็นกลุ่ม (cluster) จำนวน M กลุ่ม แล้วทำการสุ่มเลือกกลุ่มของหน่วยตัวอย่างมา m กลุ่ม โดยใช้การเลือกตัวอย่างวิธีใดวิธีหนึ่ง

การเก็บรวบรวมข้อมูลจะทำการเก็บรวบรวมจากหน่วยตัวอย่างทุกหน่วยในกลุ่มที่ถูกเลือกมาเป็นตัวอย่าง

Cluster	Population unit
Census block	Dwelling
Dwelling	Person
Day	Hour
School	Student
Employer	Employee

แผนการสุ่มตัวอย่างแบบแบ่งกลุ่ม (cluster sampling)



แผนการสุ่มตัวอย่างแบบแบ่งกลุ่ม (cluster sampling)

เมื่อไรที่จะใช้ ?

- เมื่อไม่สามารถจัดเตรียมกรอบตัวอย่างของหน่วยตัวอย่างขั้นสุดท้ายได้ หรือใช้ค่าใช้จ่ายสูงในการเตรียมกรอบตัวอย่างฯ แต่กรอบตัวอย่างหรือรายชื่อของ **cluster** สามารถหาได้สะดวก
- เมื่อค่าใช้จ่ายในการสำรวจเพิ่มขึ้น เนื่องจากการระยะทางของการเดินทางไปแต่ละหน่วยต้องคิดแยกจากกัน แต่ถ้าเป็น **cluster** หน่วยตัวอย่างภายใน **cluster** อยู่ใกล้กัน จึงทำให้ลดค่าใช้จ่ายในการเดินทางลง

ข้อเสีย : การเลือก **cluster** จำนวนหนึ่งขึ้นมาเป็นตัวอย่าง จะต้องเป็นตัวแทนของ **cluster** ที่ไม่ถูกเลือกขึ้นมา

The principles for cluster sampling:

- ควรจะมี **clusters** ขนาดเล็กจำนวนมาก ดีกว่ามี **cluster** ใหญ่ไม่กี่ **clusters**
- **Stratification and clustering** โดยปกติจะใช้ร่วมกัน

แผนการสุ่มตัวอย่างหลายขั้น (multi stage sampling)

เป็นการสุ่มหลายขั้น โดยสุ่มหน่วยใหญ่ก่อนแล้วจึงสุ่มหน่วยย่อยในหน่วยใหญ่นั้น เรียกว่า เป็นการ **sub-sample** ในทางปฏิบัติจะสุ่มอย่างมากเพียง 3 ขั้น

จุดประสงค์ของการสุ่มแบบหลายขั้น

1. เพื่อลดเวลาและความยากลำบากในการเตรียม **frame** ของหน่วยย่อยสุด
2. เพื่อประหยัดค่าใช้จ่ายในการเดินทางระหว่างหน่วยตัวอย่างในการทำงานสนาม

ข้อเสีย ของ **multi – stage** คือ ยุ่งยากทั้งในการเลือกตัวอย่างและการประมาณผล

วิธีการนี้ใช้กันแพร่หลายมากในการทำการสำรวจขนาดใหญ่

แผนการสุ่มตัวอย่างหลายขั้น (multi stage sampling)

หน่วยตัวอย่างในการสุ่มตัวอย่างแบบหลายขั้น

- หน่วยตัวอย่างขั้นที่หนึ่ง :***primary sampling units*** **or PSUs**.
- หน่วยตัวอย่างขั้นที่สอง : ***secondary sampling units*** **or SSUs**.
- หน่วยตัวอย่างขั้นสุดท้าย : ***ultimate sampling units*** **or USUs**.

แผนการสุ่มตัวอย่างแบบหลายขั้น ต้องคำนึงถึง

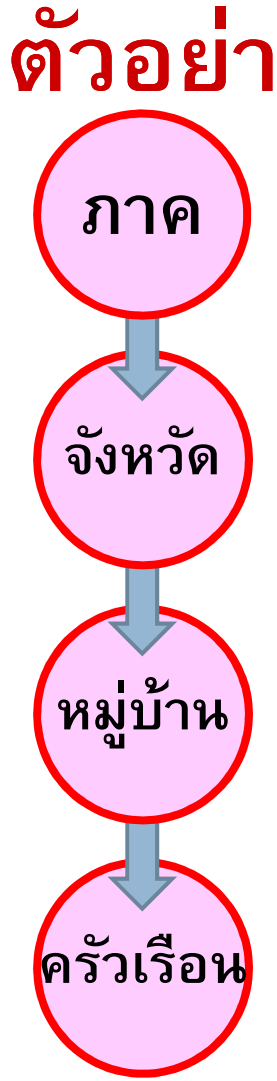
- การประมาณค่า (Estimation Domains)
- การจัดชั้นภูมิ (Stratification)
- จำนวนขั้นในการเลือกตัวอย่าง (Number of stages)
- หน่วยตัวอย่างในแต่ละขั้น (Sampling units for each stage)
- วิธีการเลือกหน่วยตัวอย่างในแต่ละขั้น (Sample selection scheme in each stage)
- กรอบตัวอย่างที่ใช้ในแต่ละขั้น *(Sampling frames used in each stage)*

แผนการสุ่มตัวอย่างหลายขั้น (multi stage sampling)

ขั้นที่ 1 เลือกจังหวัดตัวอย่าง

ขั้นที่ 2 เลือกหมู่บ้านตัวอย่าง

ขั้นที่ 3 เลือกครัวเรือนตัวอย่าง



NSO
National Statistical Office
Thailand

5. การคำนวณค่าถ่วงน้ำหนัก

ค่าถ่วงน้ำหนัก (sample weights)

- ค่าถ่วงน้ำหนัก คือ ค่าที่ระบุว่าหน่วยตัวอย่าง 1 หน่วยเป็นตัวแทนของประชากรจำนวนเท่าไหร่
- หน่วยตัวอย่างแต่ละหน่วยจะมีค่า **weight** เป็นของตัวเอง
- ในทางปฏิบัติค่า **weight** จะถูกนำไปไว้ท้าย **record** ของข้อมูลเพื่อใช้ในการประมาณค่า

ตัวอย่าง

N = 120 ราย , n = 4 ราย สอบถามอายุของแต่ละคน

ตัวอย่าง	Unweighted Value (อายุ)	ค่า Weight (เท่ากัน)	Weighted Value
นาย ก	35	30	1,050
นาย ข	65	30	1,950
นาย ค	80	30	2,400
นาย ง	45	30	1,350
Total	225	120	6,750
Mean	56.3		56.3

การคำนวณค่าถ่วงน้ำหนัก

ตัวอย่าง

N = 120 ราย , n = 4 ราย สอบถามอายุของแต่ละคน

ตัวอย่าง	Unweighted Value (อายุ)	ค่า Weight (ไม่เท่ากัน)	Weighted Value
นาย ก	35	50	1,750
นาย ข	65	20	1,300
นาย ค	80	10	800
นาย ง	45	40	1,800
Total	225	120	5,650
Mean	56.3		47.1

การคำนวณค่าถ่วงน้ำหนัก

ตัวอย่าง

- นักเรียน 100 คน ถูกสุ่มเป็นตัวอย่าง จากนักเรียนทั้งสิ้น 3,000 คน ด้วยความน่าจะเป็นที่เท่ากัน
  - ความน่าจะเป็นที่ 100 คนถูกเลือก = 100/3000
  - ดังนั้น นักเรียน 1 คน ที่ถูกเลือกจะเป็นตัวแทนของนักเรียนจำนวน 30 คน (=3000/100)

ขั้นตอนการคำนวณค่า weights

1. คำนวณ base or design weight (W_b)

W_b = ส่วนกลับของความน่าจะเป็นของหน่วยตัวอย่างที่ถูกเลือกตามแผนสุ่มตัวอย่าง

2. ปรับค่า base weight ในกรณีที่มี non-response (Adjustment for nonresponse)

$$W_{nr} = W_b * A^{nr}$$

3. ปรับค่า weight ขั้นสุดท้ายด้วยข้อมูลอื่น ๆ (Calibration Adjustment)

$$W_f = W_{nr} * A^{c}$$

1. การคำนวณ base or design weights

Base weights (W_b)

หมายถึง ส่วนกลับของความน่าจะเป็นของหน่วยตัวอย่างที่ถูกเลือก

- ค่า W_b จะขึ้นอยู่กับแผนการสุ่มตัวอย่างที่ใช้
  - Simple Random Sampling (SRS)
  - Systematic Sampling (SYS)
  - Probability Proportional to Size Sampling (PPS)
  - Stratified Sampling
  - Multi-Stage Sampling

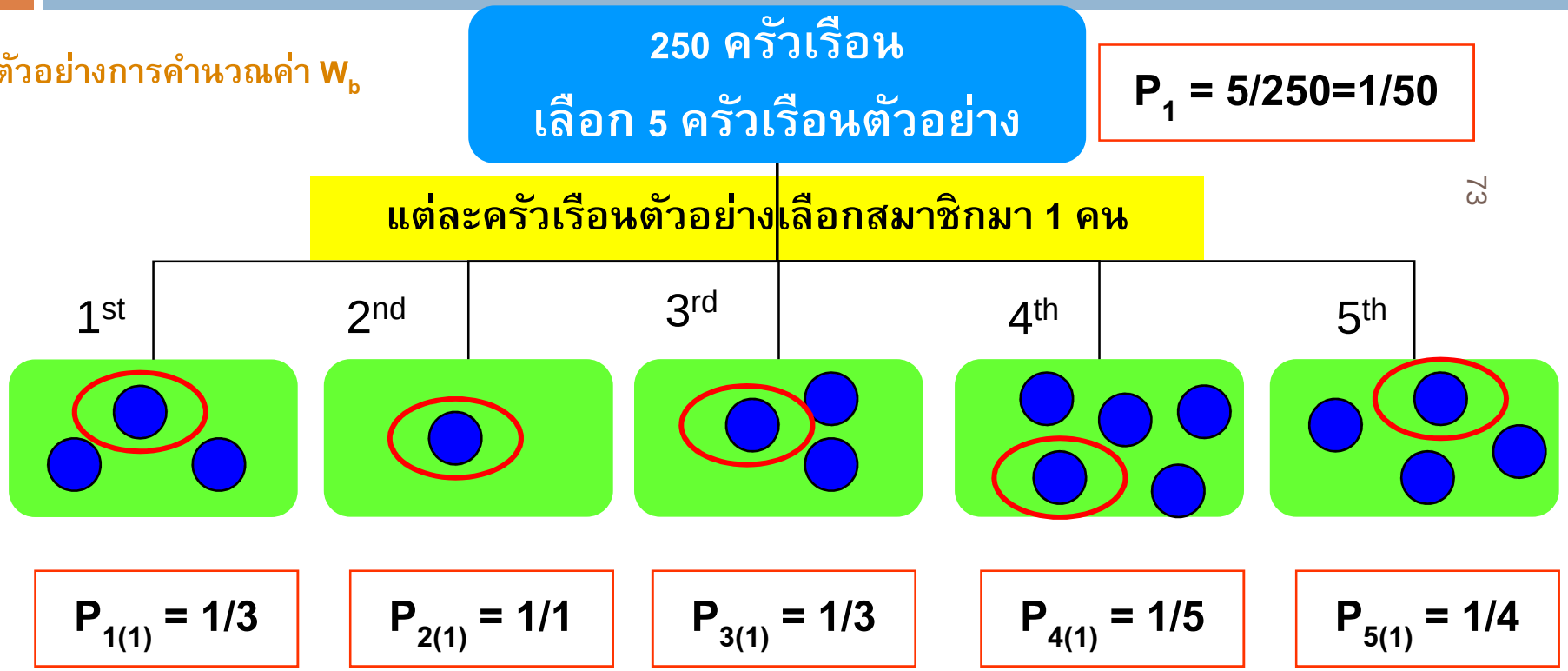
1. การคำนวณ base or design weights

ตัวอย่างการคำนวณค่า W_b

- ต้องการประมาณรายได้รวมและรายได้เฉลี่ยต่อเดือนของประชากรในจังหวัด A ซึ่งประกอบด้วย 250 ครัวเรือน โดยใช้แผนสุ่มตัวอย่างแบบ 2 ขั้น คือ
  - ขั้นแรกเลือกครัวเรือนตัวอย่าง 5 ครัวเรือน ด้วยวิธี SRS
  - ขั้นที่ 2 เลือกผู้มีรายได้ 1 คน ได้จากแต่ละครัวเรือน ด้วยวิธี SRS
- สอบถามข้อมูลเกี่ยวกับรายได้ต่อเดือน (x) และ ระดับการศึกษา (y) โดยถ้าจบปริญญาตรีขึ้นไปให้เป็น 1 และ อื่น ๆ ให้เป็น 0

1. การคำนวณ base or design weights

ตัวอย่างการคำนวณค่า W_b



$$p_{11} = \frac{1}{50} \times \frac{1}{3} = \frac{1}{150} \quad p_{12} = \frac{1}{50} \times \frac{1}{1} = \frac{1}{50} \quad p_{13} = \frac{1}{50} \times \frac{1}{3} = \frac{1}{150} \quad p_{14} = \frac{1}{50} \times \frac{1}{5} = \frac{1}{250} \quad p_{15} = \frac{1}{50} \times \frac{1}{4} = \frac{1}{200}$$

$$W_1 = \frac{1}{p_{11}} = 150 \quad w_2 = \frac{1}{p_{12}} = 50 \quad W_3 = \frac{1}{p_{13}} = 150 \quad w_4 = \frac{1}{p_{14}} = 250 \quad W_5 = \frac{1}{p_{15}} = 200$$

1. การคำนวณ base or design weights

Sampled HH	M_i	W_{bi}	x_{ij}	y_{ij}	$w_{bi}x_{ij}$	$w_{bi}y_{ij}$	$w_{bi}y_{ij}x_{ij}$
1	3	150	70	1	10,500	150	10,500
2	1	50	30	0	1,500	0	0
3	3	150	90	1	13,500	150	13,500
4	5	250	50	1	12,500	250	12,500
5	4	200	60	0	12,000	0	0
รวม	16	800	300	3	50,000	550	36,500

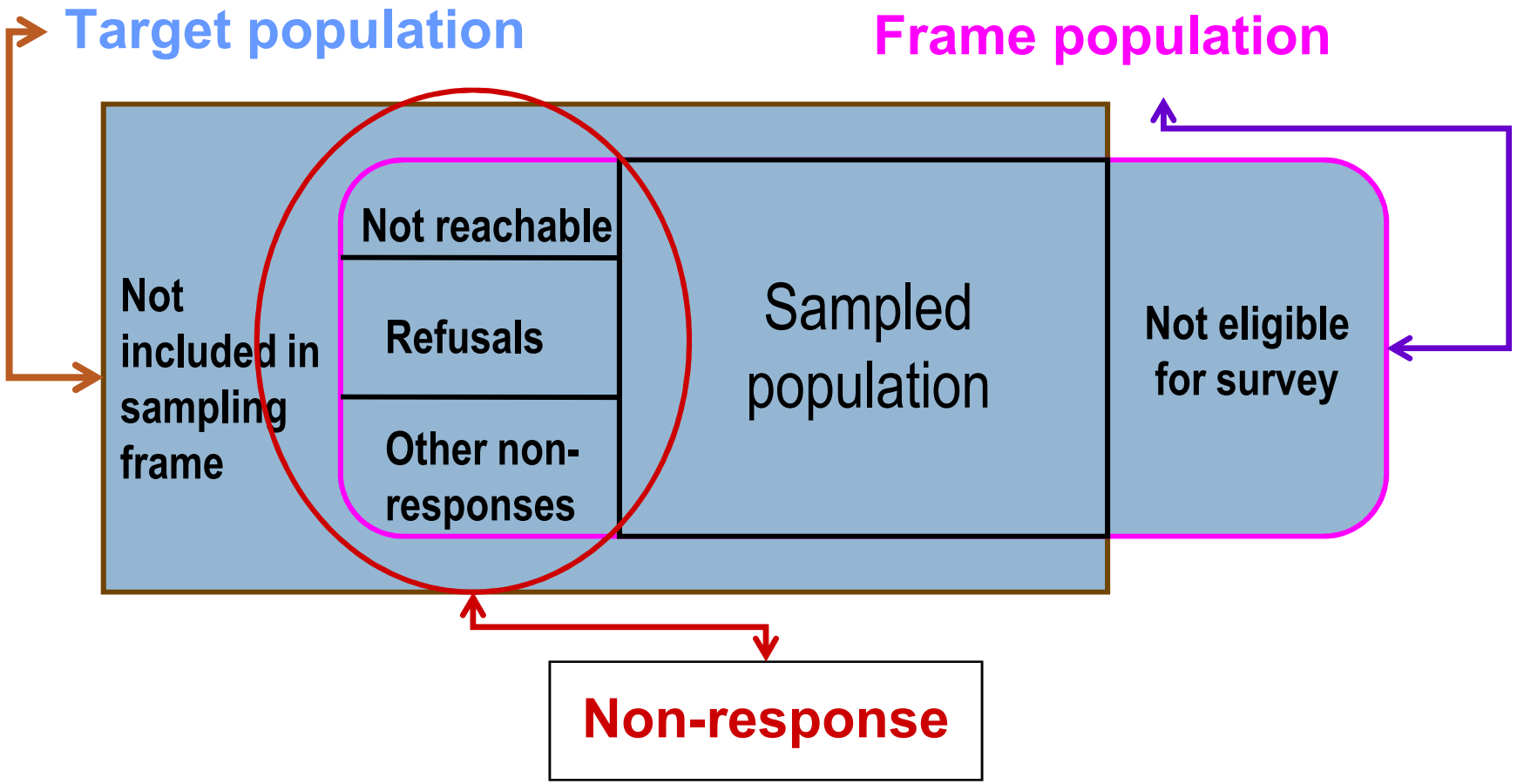
การคำนวณค่าถ่วงน้ำหนัก

1. การคำนวณ base or design weights

ค่าประมาณ	Weighted	Un-weighted
รายได้เฉลี่ย	$\bar{x}_w = \frac{\sum w_i x_{ij}}{\sum w_i} = \frac{50,000}{800} = 62.5$	300/5 = 60
สัดส่วนผู้ที่จบ ป.ตรีขึ้นไป	$\bar{x}_w = \frac{\sum w_i y_{ij}}{\sum w_i} = \frac{550}{800} = 0.6875 = 68.75\%$	3/5 = 0.6 = 60%
รายได้เฉลี่ยของผู้ที่จบ ป.ตรีขึ้นไป	$\bar{y}_w = \frac{\sum w_i y_{ij} x_{ij}}{\sum w_i y_{ij}} = \frac{36,500}{550} = 66.36$	210/3 = 70

การคำนวณค่าถ่วงน้ำหนัก

2. การปรับค่า weights ในกรณีที่มี non-response



2. การปรับค่า weights ในกรณีที่มี non-response

Non-response

- แบ่งเป็น 2 ประเภท
  - **Unit non-response** : ไม่มีข้อมูลของหน่วยตัวอย่างนั้นเลย
  - **Item non-response** : มีข้อมูลบางส่วนขาดหายไป
- การปรับค่า W_b จะปรับด้วย **Unit non-response** เท่านั้น

2. การปรับค่า weights ในกรณีที่มี non-response

- ค่านี้จะใช้เมื่อมี non-response เกิดขึ้น

$$A^{nr} = \frac{n}{n'}$$

เมื่อ n คือ ขนาดตัวอย่างที่กำหนด

n' คือ ขนาดตัวอย่างที่ได้รับจริง

(หรือบางครั้งจะใช้ sum based weights แทน)

2. การปรับค่า weights ในกรณีที่มี non-response

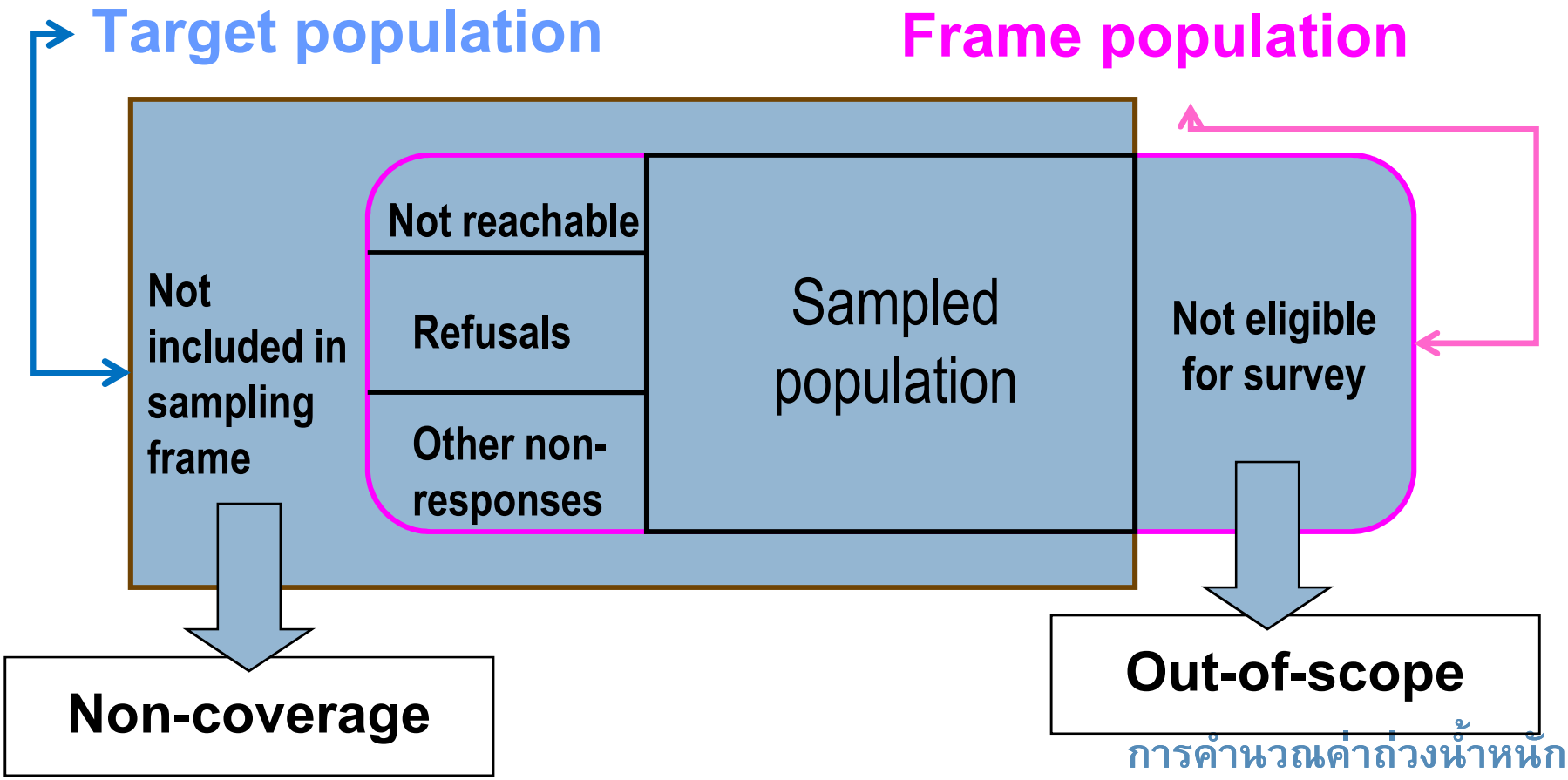
การนำค่า A^{nr} ไปใช้งาน

- นำค่า A^{nr} ไปคูณกับ $W_b \rightarrow W_{nr}$

$$W_{nr} = W_b * A^{nr}$$

3. การปรับค่า Weights ด้วยข้อมูลอื่น (Calibration Adjustment)

Coverage error



การคำนวณค่าถ่วงน้ำหนัก

3. การปรับค่า Weights ด้วยข้อมูลอื่น (Calibration Adjustment)

Coverage error

- เกิดจากความไม่สมบูรณ์/ไม่ทันสมัยของ Sampling Frame
- การแก้ปัญหา
  - พัฒนาการทำงานสนามให้มีประสิทธิภาพมากขึ้น เช่น พัฒนากระบวนการจัดทำ listing
  - ใช้ sampling frame จากหลายแหล่ง
  - ปรับค่า Weight ด้วย calibration factors เช่น ปรับค่ายอดรวมประชากรด้วยค่า projection หรือ ใช้ข้อมูลจากทะเบียนราษฎร์ เป็นต้น

3. การปรับค่า Weights ด้วยข้อมูลอื่น (Calibration Adjustment)

$$A^c = \frac{N^c}{N^e}$$

เมื่อ N^c คือ จำนวนประชากรจากข้อมูลแหล่งอื่น

N^e คือ จำนวนประชากรที่ได้จากการประมาณค่า

Final weights = $W_f = W_{nr} * A^c$