

บทที่ 2

การวางแผนดำเนินการทางสถิติสำหรับการสำมะโน และการสำรวจด้วยตัวอย่าง

ขั้นของการดำเนินงานทางสถิติ

กระบวนการรวบรวมข้อมูลสถิติ เราอาจจะแบ่งออกเป็นขั้นใหญ่ ๆ ได้ 4 ขั้น

ดังนี้ :-

1. ขั้นวางแผนและเตรียมงาน
2. ขั้นงานสนาม ได้แก่ การเก็บรวบรวมข้อมูลปฐม
3. การประมวลผลข้อมูล ได้แก่ การผลิตข้อมูลสถิติที่ต้องการ จากข้อมูลปฐมที่รวบรวมมาได้
4. ขั้นการวิเคราะห์ เพื่อตอบความถูกต้องของข้อมูลและการตีความหมาย นำข้อมูลสถิติมาใช้ให้เป็นประโยชน์

ทั้งนี้เราจะไม่เน้นการเก็บรวบรวมสถิติจากแหล่งทุติยภูมิ

แต่ละขั้นใหญ่ของการดำเนินงานทางสถิติ ยังสามารถแยกรายละเอียดออกไป

ได้อีกหลายขั้น ดังนี้ :

ขั้นวางแผนและเตรียมงาน

การกำหนดข้อมูลสถิติที่ต้องการในรายละเอียด

จากเป้าหมายของการใช้สถิติและความต้องการในข้อมูลสถิติอย่างคร่าว ๆ ผู้เตรียมงานจะต้องเตรียมตารางสถิติที่ต้องการทุกตาราง และแสดงแบบฟอร์มของตาราง (table format) แสดงจำนวนสดมภ์ของตาราง แสดงรายละเอียดของการจัดจำแนกในสดมภ์ (column caption) และจำนวนบรรทัด (row) พร้อมทั้งรายละเอียดของการจัดจำแนกในแต่ละบรรทัด (stub items) จะต้องดูว่าทั้งตารางมีข้อมูลสถิติทั้งหมดกี่ยอด แต่ละยอดสถิติจะต้องการตัวเลขกี่ตัว เพราะค่าใช้ง่ายในการเก็บข้อมูลสถิตินั้น ขึ้นกับจำนวนตารางและจำนวนยอดสถิติโดยตรง แต่ละตารางจะมีชื่อหัวตาราง (table title) ซึ่งบรรยายข้อมูลในตารางได้อย่างถูกต้อง ซึ่งในบางครั้งนอกจากหัวตารางแล้วยังต้องมีหัวตารางย่อยอีกด้วย

ตามปกติจะต้องมีการปรึกษาหารือกันอย่างใกล้ชิดระหว่างผู้ดำเนินการเก็บสถิติ กับผู้ที่จะต้อง ใช้สถิติโดยตรง ในกรณีที่มีการเก็บสถิติเป็นสถิติในระดับชาติ เช่นการเก็บสถิติ

เกี่ยวกับประชากรในโครงการสำมะโนประชากร หน่วยงานผู้เก็บจะต้องปรึกษากับหน่วยงานผู้ใช้หลายๆ หน่วยงาน ซึ่งอาจจะเป็นทั้งหน่วยงานของรัฐและของ เอกชน ในกรณีเช่นนี้หลังจากที่ปรึกษากันเสร็จแล้ว ข้อมูลสถิติที่รวบรวมมาได้อาจจะไม่ถูกต้องตามประสงค์ของหน่วยงานใดหน่วยงานหนึ่งโดยตรง บางครั้งอาจจะมีมากเกินไปความต้องการของหน่วยงานบางหน่วย อาจจะไม่ค่อยไปสำหรับบางหน่วย ผู้ที่หน้าที่เก็บสถิติมักจะ ได้ยินคำตำหนิเป็นครั้งคราวว่า "ข้อมูลที่เก็บไม่เห็นจะนำมาใช้เป็นประโยชน์" หรือบางครั้งก็มีผู้ตำหนิว่า "ข้อมูลมีมากเกินไปหมดจนไม่ทราบว่าจะเอาส่วนไหนมาใช้"

หน่วยงานที่เล็งลงไปเช่น บริษัทห้างร้านที่จำเป็นต้องใช้สถิติเกี่ยวกับการขาย วิจัยตลาด มักจะมีความต้องการในข้อมูลสถิติเพื่อใช้เฉพาะเรื่องเป็นการแน่นอน นักสถิติผู้ทำหน้าที่เก็บสถิติก่อน จะดำเนินงานจะต้องแสดงตารางสถิติ (table output) ให้ผู้ใช้ได้ดูเสียก่อนต่อเมื่อเห็นชอบแล้วจึงจะดำเนินการต่อไป

ถึงกระนั้น ก็มีอยู่บ่อยครั้งที่ผู้ใช้สถิติเองขาดประสบการณ์เกี่ยวกับการใช้สถิติ และขาด statistical maturity ไม่ทราบแน่ชัดว่าสถิติอะไรเกี่ยวข้องกับการบริหารงาน หรือวางแผนงานของตน ปล่อยให้ให้นักสถิติผู้รวบรวมสถิติดำเนินการไปเอง เสียเงินไปเป็นจำนวนมาก มาทราบเอาเมื่อเก็บข้อมูลมาแล้ว (หลังจากรอคอยเป็นเวลาหลายเดือนหรือบางครั้งเป็นปี) ว่าสถิติดังกล่าวหาใช้สถิติที่ตนต้องการไม่ ทั้งไว้ให้นักสถิติดำเนินการเอง นักสถิติมักจะมีแนวโน้มเอียงที่เก็บตัวเลขมากเกินไป จนเกินต้องการจึงสิ้นเปลืองมาก และเมื่อมีตัวเลขมากเกินไปต้องการก็ยากแก่การวินิจฉัยว่า ข้อมูลรายการไหนจะนำมาใช้ให้เป็นประโยชน์ที่สุด การกำหนดข้อมูลสถิติที่ต้องการจึงต้องอาศัยความคิดอย่างรอบคอบก่อนที่จะดำเนินการทางสถิติ

การกำหนดวิธีดำเนินการ

ข้อมูลสถิติที่ต้องการจะรวบรวมอาจจะทำได้โดยวิธีการหลัก 4 วิธีด้วยกัน คือ

- 1) โดยวิธีสำมะโน
- 2) โดยวิธีสำรวจด้วยตัวอย่าง
- 3) โดยรายงานจากหน่วยบริหาร
- 4) โดยวิธีการทะเบียน

วิธีสำมะโน

การเก็บข้อมูลจากแหล่งปฐม ตามปกติเราจะต้องกำหนดลงไปให้แน่นอนว่า เราจะใช้อะไรเป็นแหล่งข้อมูลปฐม แหล่งของข้อมูลปฐมนี้เราเรียกว่า หน่วยแฉงนับ (enumeration unit) การออกไปเก็บข้อมูลเบื้องต้นจากหน่วยแฉงนับเราเรียกว่า แฉงนับ (enumerate)

นอกจากนั้นหลังจากกำหนดหน่วยแฉงนับแล้ว ยังต้องกำหนดลงไปด้วยว่า ขอบข่ายของการเก็บข้อมูลสถิตินั้น จะให้ครอบคลุมถึงหน่วยแฉงนับไหน ในเนื้อที่ใดบ้าง ขอบข่ายดังกล่าวเราเรียกว่า คุ่มรวม ของการเก็บข้อมูล และหน่วยแฉงนับทั้งหมดที่อยู่ในคุ่มรวม เราจะเรียกว่า จักรวาล (universe)

วิธีของการสำมะโนกำหนดไว้ว่า ข้อมูลสถิติที่ประมวลได้จะต้องได้มาจากการแฉงนับหน่วยแฉงนับทุกหน่วยในคุ่มรวม

เช่น ในสำมะโนประชากร 2513 เรากำหนดให้ "ครัวเรือน" เป็นหน่วยแฉงนับ และมีคุ่มรวมไปถึงทุกครัวเรือนในประเทศไทย นอกจากสถานทูตของชาวต่างประเทศ และครัวเรือนของชนชาวเขาเผ่าต่าง ๆ ซึ่งอยู่นอกสารบบการบริหารของราชการไทย นอกจากนี้คุ่มรวมยังรวมถึงครัวเรือนคนไทยที่ทำงานตามสถานทูตไทยในต่างประเทศอีกด้วย

คำว่าสำมะโนนี้ตรงกับภาษาอังกฤษว่า census แต่เดิมมาคำว่า census หมายถึงสำมะโนครัว ซึ่งได้แก่การนับบ้านที่อยู่อาศัยและพลเมืองเท่านั้น ไม่เกี่ยวกับการนับสัตว์และสิ่งของ รากศัพท์เดิมของคำนี้มาจากคำว่า Censere ในภาษาละตินแปลว่า ตีราคาหรือเก็บภาษี ในสมัยจักรวรรดิโรมันหน้าที่อย่างหนึ่งของ censor ก็คือจัดทำทะเบียนจำนวนและทรัพย์สินสมบัติของราษฎร เพื่อตีราคาและเรียกเก็บภาษี

สำหรับประเทศไทยประชาชนส่วนมากคุ้นกับคำว่าสำมะโนครัว ซึ่งหมายถึงการนับบ้านนับครัวเรือนและประชากร ซึ่งในประเทศไทยได้เคยทำมาแล้ว 9 ครั้ง 5 ครั้งแรกอยู่ในความรับผิดชอบของกระทรวงมหาดไทย (คือทำในปี 2453, 2462, 2472, 2480 และ 2490) ครั้งที่ 6 ถึง 9 ทำในปี 2503, 2513, 2523, 2533 ได้เปลี่ยนชื่อมาเรียกว่าสำมะโนประชากรและอยู่ในความรับผิดชอบของสำนักงานสถิติแห่งชาติ

ในระยะหลังเราให้ความหมายของคำสำมะโนกว้างออกไป โดยที่สำมะโนไม่จำเป็นจะต้องหมายถึงการเก็บสถิติเกี่ยวกับประชากรเท่านั้น ได้มีการทำสำมะโนเกษตร

สำมะโนอุตสาหกรรม สำมะโนธุรกิจการค้าและบริการ สำมะโนการประมง เป็นต้น ไม่ว่าจะเก็บสถิติในเรื่องใด หากมีการแจกจ่ายหน่วยแจกจ่ายในคัมรวม เราจะจัดอยู่ในประเภทสำมะโนทั้งสิ้น (ดูนิยามในพระราชบัญญัติสถิติ 2508)

วิธีสำรวจด้วยตัวอย่าง (sample survey)

คือการเก็บข้อมูลสถิติโดยการประมาณยอดข้อมูลสถิติที่ต้องการของคัมรวมจากข้อมูลปฐมที่รวบรวมมาได้จากหน่วยแจกจ่าย "ตัวอย่าง" เพียงบางหน่วย หน่วยแจกจ่ายและคัมรวมก็ยังคงต้องกำหนดให้มีขึ้นเช่นเดียวกับสำมะโน

การสำรวจด้วยตัวอย่าง จะต้องทำการกำหนดขนาดของตัวอย่าง และมีการสุ่มหน่วยแจกจ่ายตัวอย่าง ซึ่งจะเป็นตัวแทนของหน่วยแจกจ่ายที่มีทั้งหมดในคัมรวม การแจกจ่ายจะทำเฉพาะที่หน่วยแจกจ่ายตัวอย่างเท่านั้น และหลังจากเก็บข้อมูลเบื้องต้นได้แล้ว การประมวลยอดสถิติ จะต้องเอาข้อมูลเบื้องต้นมาเข้าสู่การประมาณจึงจะเป็นยอดข้อมูลสถิติของคัมรวมที่ต้องการ

วิธีรายงานจากหน่วยบริหาร

ข้อมูลปฐมบางประเภทเกิดขึ้นมาเป็นผลพลอยได้จากการบริหารราชการหรือการบริหารหน่วยงาน ทำให้สามารถประมวลเป็นยอดข้อมูลสถิติได้ ในกรณีนี้ข้อมูลสถิติที่จะประมวลได้จะถูกกำหนดโดยข้อมูลที่เป็นผลพลอยได้จากการบริหารนั้น ๆ เช่น ใบนำส่งสินค้า ซึ่งจะต้องผ่านมือเจ้าหน้าที่ศุลกากรในการนำสินค้าเข้า/ออก จะเป็นแหล่งของข้อมูลปฐมสามารถทำให้ประมวลยอดข้อมูลสถิติแสดงปริมาณการค้าระหว่างประเทศได้

หรือใบแจ้งย่นภาษี ภงด. ซึ่งผู้ประเมินจะต้องยื่นต่อกรมสรรพากร อาจจะนำมาใช้ประมวลข้อมูลสถิติเกี่ยวกับการเก็บภาษี

รายงานผลการปฏิบัติงานของโรงเรียนภายในสังกัดของกรมในกระทรวงศึกษาธิการ ก็นำมาใช้ในการประมวลสถิติที่จำเป็นต่อการบริหารของกรมนั้น

ในบางครั้งเราอาจวางแผนรวบรวมข้อมูลสถิติโดยอาศัยองค์การหรือหน่วยบริหารของรัฐที่มีอยู่แล้ว และสถิติดังกล่าวไม่จำเป็นต้องเป็นผลพลอยได้จากการบริหารงานของหน่วยงานนั้น ๆ แต่เป็นผลพลอยได้ของหน่วยบริหารที่มีอยู่แล้ว

หน่วยบริหารของรัฐที่เป็นแหล่งกำเนิดของข้อมูลปฐมมากที่สุดในประเทศไทยเห็น

จะได้แก่ ผู้ใหญ่บ้าน โครงการเก็บข้อมูลสถิติจากรายงานของผู้ใหญ่บ้านมีอาทิ เช่น รายงานการเกษตร¹ มีรายละเอียดเนื้อหาเฉพาะปลูกข้าว (และพืชอื่น ๆ) เนื้อที่เก็บเกี่ยว เนื้อที่เสียหาย และปริมาณผลผลิตที่เก็บได้ภายในหมู่บ้านในสังกัด

การรวบรวมสถิติจากรายงานระดับหมู่บ้านทุกหมู่บ้านในประเทศไทย มองในทัศนะหนึ่งอาจจะจัดได้ว่าเป็นการสำมะโนประเภทหนึ่ง โดยมีผู้ใหญ่บ้านเป็นหน่วยแจงนับหรือมองอีกทัศนะหนึ่งเป็นการประมวลยอดข้อมูลจากข้อมูลระดับหมู่บ้าน ซึ่งเป็นข้อมูลทุติยภูมิกำลังกันอยู่ก็ได้ แต่โครงการสถิติดังกล่าวมีลักษณะดำเนินงานเพื่อการสถิติโดยเฉพาะ ไม่ใช่เป็นสถิติซึ่งเกิดจากผลพลอยได้ในการบริหาร แต่เป็นสถิติที่ได้โดยอาศัยมือของหน่วยงานบริหารที่มีอยู่แล้ว

วิธีการเก็บสถิติดังกล่าว จากองค์การบริหารของรัฐ ไม่ว่าจะ เป็นผลพลอยได้ในการบริหารหรือไม่เป็นก็ตาม จัดเข้าอยู่ในประเภทวิธีรวบรวมสถิติจากรายงานทั้งสิ้น

วิธีรวบรวมสถิติจากระบบทะเบียน

การทะเบียนคือการเก็บข้อมูลบางประเภทอย่างต่อเนื่องกันไป โดยมีการปรับและแก้ไขข้อมูลบางรายการให้ทันสมัย ตามรายการเปลี่ยนแปลงของข้อมูลนั้น ๆ

ข้อมูลที่เก็บในการทะเบียนมีลักษณะเป็นข้อมูลเบื้องต้น การนำข้อมูลมาใช้ก็นำมาใช้เป็นราย ๆ ไปเพื่อประโยชน์ในการบริหาร และมักจะมีพระราชบัญญัติให้อำนาจในการเก็บหรือดำเนินการเกี่ยวกับการทะเบียนนั้น ๆ

ตัวอย่างการทะเบียนที่มีการจัดทำเป็นระบบข้อมูลใหญ่ที่สุด และเก่าแก่ที่สุดในประเทศไทย ได้แก่การทะเบียนราษฎร (จัดทำโดย พรบ. การทะเบียนราษฎร พ.ศ. 2499) ซึ่งมอบอำนาจให้กระทรวงมหาดไทย (ได้แก่เทศบาลสำหรับส่วนที่อยู่ในเขตเทศบาล และอำเภอในส่วนที่อยู่นอกเขตเทศบาล) จัดทำทะเบียนบ้าน ซึ่งมีชื่อของบุคคลแต่ละคนที่อาศัยอยู่ในบ้าน และข้อมูลบางรายการของแต่ละบุคคลในบ้าน ข้อมูลเหล่านี้มี เพศ สถานภาพการสมรส อายุ ระดับการศึกษา และหมายเหตุแสดงข้อเท็จจริงอื่น ๆ ของบุคคล ตามแต่นายทะเบียนจะเห็นเหมาะสม ในระบบทะเบียนได้มีการกำหนดไว้ ให้แก้ไขทะเบียน

¹ การประมวลยอดข้อมูลสถิติ ทำโดยเจ้าหน้าที่ส่วนกลาง กระทรวงเกษตร

ราษฎรให้ถูกต้องตามความเป็นจริง ในกรณีที่มีการเปลี่ยนแปลงเกิดขึ้นในข้อมูลรายการของแต่ละบุคคลในบ้าน หรือเมื่อมีบุคคลเกิดขึ้นใหม่ในบ้าน หรือบุคคลย้ายเข้าหรือย้ายออกไปจากบ้าน เป้าหมายหลักของการทะเบียนราษฎรก็เพื่อประโยชน์ของการปกครองเป็นประการสำคัญประการแรกเจ้าหน้าที่ปกครองสามารถติดตามความเคลื่อนไหวของพลเมืองในประเทศ มีหลักฐานยืนยันบุคคล และภูมิลำเนาที่อยู่ประจำของแต่ละบุคคลในกรณีที่มีความจำเป็นติดต่อทางราชการ เช่น มาขอใบขับขี่ ยานพาหนะ ทำบัญชีเลือกตั้ง บัญชีผู้อยู่ในวัยเกณฑ์ทหาร หรือบัญชีผู้เข้าเกณฑ์การศึกษาระดับประถมศึกษา เป็นต้น

โดยทั่วไป การจัดการกระทำตามระเบียบวิธีเชิงสถิติเพิ่มเติมกับข้อมูลเบื้องต้นที่เก็บไว้ในทะเบียนอาจทำให้เกิดข้อมูลสถิติที่เป็นประโยชน์ได้ เช่น จากระบบทะเบียนราษฎรก็สามารถประมวลยอดจำนวนประชากรเป็นรายตำบล รายอำเภอ และรายจังหวัด ยอดจำนวนคนเกิด/ตายภายในคาบเวลาหนึ่ง ๆ ยอดจำนวนคนย้ายเข้าอำเภอ และย้ายออกจากอำเภอภายในคาบเวลาหนึ่ง ๆ เป็นต้น

การจัดทะเบียนยานพาหนะของกรมตำรวจ ก็ทำให้สามารถประมวลยอดจำนวนยานพาหนะจำแนกตามประเภทต่าง ๆ ภายในแต่ละจังหวัด

หรือทะเบียนข้าราชการพลเรือน (กพ.7) ซึ่งเก็บอยู่ ณ สำนักงาน ก.พ. แสดงรายการ อายุ วุฒิ ตำแหน่ง กรม กอง สถานภาพสมรส เงินเดือน ฯลฯ ของข้าราชการแต่ละคนก็อาจจะนำมาใช้เป็นแหล่งเกิดของข้อมูลสถิติเกี่ยวกับข้าราชการพลเรือน เช่น จำนวนข้าราชการเป็นรายกรมจำแนกตามวุฒิต่าง ๆ จำนวนข้าราชการจำแนกเป็นรายอายุ เป็นต้น

ข้อควรพิจารณาในการกำหนดวิธีดำเนินการ

การรวบรวมข้อมูลสถิติจากวิธีการทั้ง 4 วิธี ดังกล่าวมาแล้วมีข้อแตกต่างกันอย่างมาก ทั้งในเรื่องเกี่ยวกับวิธีการดำเนินงานทางสถิติ ในเรื่องคุณภาพของข้อมูล และในเรื่องเกี่ยวกับงบประมาณที่จะใช้ในการรวบรวมข้อมูลสถิติ

ในบางครั้งข้อมูลสถิติรายการเดียวอาจจะรวบรวมจากวิธีการมากกว่า 1 วิธี ข้อมูลที่รวบรวมได้จากวิธีการต่าง ๆ กัน ซึ่งดูเผิน ๆ อาจจะเป็นข้อมูลที่ซ้ำรายการกัน แต่หากผู้รู้ซึ่งถึงแหล่งที่มา วิธีการและคุณภาพของข้อมูลแล้ว ข้อมูลรายการเดียวกันและจากวิธีการต่างกันอาจจะนำมาสนับสนุน (supplement) กันได้

งานสำมะโนเป็นงานที่จัดดำเนินการโดยรัฐ เป็นงานค่อนข้างใหญ่ และต้องสิ้นเปลืองงบประมาณเป็นจำนวนมาก ทั้งนี้เพราะเป็นงานที่มีผู้รวมทั้งประเทศมีหน่วยเจ้านับเป็นจำนวนมากที่จะต้องเจ้านับ แต่ละโครงการใช้งบประมาณ ประมาณ 20-60 ล้านบาท โครงการสำมะโนเท่าที่จัดทำแล้วในประเทศไทยมีดังนี้

	ทำครั้ง สุดท้าย	ความถี่	หน่วยเจ้านับเมื่อ สำมะโนครั้งสุดท้าย	งบประมาณโดย ประมาณครั้ง สุดท้าย	ทำมา แล้ว
<u>สำมะโนประชากร</u>	2533	ทุก 10 ปี	ประมาณ 12.3 ล้าน ครัวเรือน	246 ล้านบาท	9 ครั้ง
<u>สำมะโนเกษตร</u>	2536	-	ประมาณ 5.9 ล้าน ครัวเรือนเกษตร	362 ล้านบาท	4 ครั้ง
<u>สำมะโนอุตสาหกรรม</u>	2507	-	160,000 สถาน ประกอบการ	15 ล้านบาท	1 ครั้ง
<u>สำมะโนธุรกิจ</u> (เฉพาะในเขตเทศบาล)	2531	-	500,000 สถาน ประกอบการ	88.9 ล้านบาท	2 ครั้ง
<u>สำมะโนประมง</u>	2528	-	74,494 ครัวเรือนประมง	30 ล้านบาท	2 ครั้ง

ในทางปฏิบัติประเทศต่าง ๆ จะไม่ทำสำมะโนในเรื่องหนึ่ง ๆ ซ้ำทุกปีเพราะปัญหาเรื่องงบประมาณและกำลังคนที่จะต้องใช้ในการทำงานประการหนึ่ง และนอกจากนั้นงานสำมะโนแต่ละงานมีคาบระยะเวลาการทำงานยาวนาน 2-3 ปี งานจึงจะแล้วเสร็จทั้งโครงการ ประเทศที่เจริญแล้วที่มีทรัพยากรมาก และมีความจำเป็นในการใช้สถิติอาจจะทำสำมะโนในเรื่องหนึ่ง ๆ ถี่กว่าประเทศที่ด้อยพัฒนา เช่น ประเทศญี่ปุ่น ทำสำมะโนประชากรทุก ๆ 5 ปี คือ ปี ค.ศ. ที่ลงท้ายด้วย 5 และ 0 ประเทศไทยเราทำทุก ๆ 10 ปี สำมะโนอื่นก็เช่นกัน ปกติแล้วจะทำด้วยความถี่ 10 ปี หรือ 5 ปีครั้ง บางประเทศที่ขาดทรัพยากรมาก ๆ อาจจะทำสำมะโนด้วยความถี่น้อยกว่านี้ก็ได้ หรือบางประเทศในบางเรื่องจนบัดนี้ยังไม่เคยทำสำมะโนเลย

ข้อดีเด่นของสำมะโนก็คือ ยอดข้อมูลสถิติที่รวบรวมได้ในทุกเรื่องสามารถแสดงออกในเขตบริหารหรือในเขตภูมิศาสตร์ที่เล็กที่สุดได้ ทั้งนี้เพราะมีข้อมูลเบื้องต้นจากทุกหน่วยแรงแบบโน้มนำเมื่อจะรวมเป็นยอดข้อมูลสถิติของเขตบริหารได้ก็ย่อมจะทำได้ เช่นในสำมะโนเกษตรเราอาจจะแสดงตารางข้อมูลสถิติทุกตารางในระดับจังหวัด ระดับอำเภอ ระดับตำบล หรือแม้ในระดับหมู่บ้านก็ย่อมจะทำได้

แต่ถึงกระนั้นสำมะโนก็ยังมีข้อเสียอยู่มาก ประการแรก เนื่องจากต้องใช้ทรัพยากรมาก ไม่สามารถจะทำงานสำมะโนได้ทุกปี ทำให้ข้อมูลไม่ครบทุกปีอาจไม่เพียงพอกับการใช้ ประการที่สอง เนื่องจากมีหน่วยแรงแบบโน้มนำเป็นจำนวนมากทำให้มีปริมาณงานมากเป็นเงาตามตัว งานสำมะโนจะไม่สามารถทำให้แล้วเสร็จโดยทันที่ ข้อมูลผลิดอกมาในการทำสำมะโนแต่ละครั้งอาจจะทำเสร็จออกมา 2-3 ปีหลังจากวันเริ่มแรงแบบโน้มนำ ความจริงการปฏิบัติงานของสำมะโนบางครั้งจะยาวนานกว่า 2-3 ปี เสียอีก เพราะจะต้องใช้เวลาเตรียมงาน และทดสอบเทคนิคการทำงานประมาณ 2 ปีก่อนทำการแรงแบบโน้มนำใหญ่ ๆ แต่ละครั้งจะมีคาบการทำงานยาวนานร่วม 5 ปี ประการที่สาม ข้อมูลที่รวบรวมได้จากสำมะโนก็ยังมีคุณภาพเป็นที่น่าสงสัยอยู่มาก เพราะต้องใช้เจ้าหน้าที่ร่วมทำสำมะโนเป็นจำนวนมากประการหนึ่ง เจ้าหน้าที่เหล่านี้ (โดยเฉพาะพนักงานแรงแบบโน้มนำ) ไม่ใช้พนักงานสำมะโนอาชีพ ส่วนมากเป็นผู้มีอาชีพอื่น (นักศึกษาบ้าง ครูบ้าง) ต้องมารับการอบรมเกี่ยวกับงานอย่างเร่งรีบ (ประมาณ 1-2 อาทิตย์) และเมื่อรับการอบรมเสร็จแล้วก็ต้องเร่งทำงานให้เสร็จภายในคาบเวลาที่กำหนดให้ เช่น ในสำมะโนเกษตร หากถามถึงจำนวนเนื้อที่เพาะปลูกและผลผลิตของพืชแต่ละประเภท ย่อมไม่อยู่ในวิสัยของกลีกรจะตอบให้ถูกต้องทุกคนไป และเนื่องจากสถิติที่ถูกต้องจะต้องมาจากข้อมูลเบื้องต้นที่ถูกต้องทุกรายการของหน่วยแรงแบบโน้มนำโดยไม่มีข้อยกเว้น เมื่อข้อมูลเบื้องต้นจากหน่วยแรงแบบโน้มนำมีผิดพลาดเช่นนี้ ยอดข้อมูลสถิติก็ผิดไปจากความเป็นจริงด้วย เนื่องจากสำมะโนมีปริมาณงานมาก ย่อมไม่สามารถที่จะใช้เทคนิคการวัดหรือเทคนิคการสัมภาษณ์ที่พิสดารเพื่อให้ได้ข้อมูลเบื้องต้นที่ถูกต้องตามความเป็นจริงได้ ตัวอย่างที่เห็นได้ชัดในเรื่องนี้ได้แก่ สำมะโนเกษตร 2506 ข้อมูลเกี่ยวกับจำนวนข้าวโพดที่ผลิตได้ซึ่งรวมมาจากข้อมูลเบื้องต้นจากแต่ละครัวเรือนเกษตร ได้ยอดการผลิต 550,000 ตันในคาบเวลาดังแต่่มกราคม-ธันวาคม 2505 แต่ในปีพ.ศ. 2505 ประเทศไทยส่งข้าวโพดออกจำหน่ายยังต่างประเทศกว่า 1,000,000 ตัน ข้อมูลที่รวบรวมได้น้อยไปกว่าความเป็นจริงอย่างน้อยร้อยละ 45 ประการสุดท้าย

ปัญหาเรื่องการบริหารงานและควบคุมคุณภาพในงานแขนงนี้ ซึ่งเป็นปัญหาหนึ่งที่ยากที่สุดใน การทำสำมะโน ทั้งนี้เพราะมีจำนวนบุคคลที่ร่วมงานเป็นจำนวนมาก (กว่า 20,000 คน) จากที่ต่าง ๆ และจากภูมิหลังต่าง ๆ กัน เนื่องจากต้องการพนักงานเป็นจำนวนมากไม่มี โอกาสให้กลิ่นกรองรับแต่คนดีเข้ามาทำงาน และบุคคลเหล่านี้เมื่อออกปฏิบัติงานในสนาม ก็ต้องออกไปทำงานเป็นอิสระแต่ผู้เดียว ไม่มีการควบคุมอย่างใกล้ชิด พนักงานอาจจะทำงาน อย่างรวบรัด ลักแต่ให้เสร็จ หรือทุจริตไม่ทำงาน แต่กรอกตัวเลขเบื้องต้นมาให้ (ยกเมฆ) จึงเป็นการยากที่จะให้ผลงานมีคุณภาพสูงได้

การสำรวจด้วยตัวอย่าง (sample survey) หรือบางครั้งก็เรียกว่า (sam- ple census) หรือบางครั้งก็เรียกว่า (survey research) มีลักษณะดำเนินการคล้าย คลึงกับสำมะโน ผิดกันตรงที่หลังจากกำหนดข้อมูลที่จะเก็บ และกำหนดคุ่มรวมและหน่วยแจง นับแล้ว การแจงนับทำที่หน่วยแจงนับตัวอย่าง เป็นบางหน่วยเท่านั้น เนื่องจากค่าใช้จ่ายใน การเก็บสถิติส่วนใหญ่อยู่ที่การออกไปเก็บข้อมูลเบื้องต้นตอนแจงนับ เมื่อใช้วิธีการแจงนับ จากตัวอย่างเช่นนี้ จึงทำให้การสำรวจด้วยตัวอย่างได้เปรียบการสำมะโนตรงที่ทุนค่าใช้จ่าย ในการดำเนินงานเป็นอย่างมาก โครงการสำมะโนอาจจะใช้งบประมาณ 10 - 50 ล้านบาท และการสำรวจด้วยตัวอย่างใช้งบประมาณไม่กี่พันบาทขึ้นไป โครงการสำรวจที่ ใช้งบประมาณ 3 แสนบาทขึ้นไป จัดได้ว่าเป็นโครงการสำรวจที่ค่อนข้างจะใหญ่ ทั้งนี้ต้อง แล้วแต่ว่าขนาดของตัวอย่างเป็นเท่าไร จำนวนเงินงบประมาณที่จะใช้ทำสำรวนับว่าอยู่ ในวิสัยที่หน่วยงานเล็ก ๆ เช่น บริษัทห้างร้านจะสามารถจ่ายได้ หรือแม้แต่หน่วยงาน ระดับชาติ การใช้วิธีสำรวจด้วยตัวอย่างแทนสำมะโน ก็ช่วยทุ่นงบประมาณแผ่นดินไปได้ มาก เทคนิคการสำรวจด้วยตัวอย่างจึงนับได้ว่าเป็นเครื่องมือสำคัญในการเก็บข้อมูลสถิติ โดยไม่ต้องเสียค่าใช้จ่ายมากนัก สถิติระดับชาติหลายรายการได้มาโดยวิธีนี้ งานวิจัย ธุรกิจหรืองานวิจัยในสนามโดยสัมภาษณ์ทั่ว ๆ ไป ก็จัดได้ว่าเป็นการสำรวจประเภทหนึ่ง

ข้อมูลที่รวบรวมได้จากการสำรวจ จำเป็นที่จะต้องมีความคลาดเคลื่อนจากการ สุ่มตัวอย่างทุกครั้ง และมีความจำเป็นจะต้องสุ่มตัวอย่างที่เป็นตัวแทนของหน่วยแจงนับทั้ง หมดในคุ่มรวมโดยไม่ให้ความเียงเอน (bias) ซึ่งในบางครั้งอาจจะปฏิบัติได้โดยไม ่ง่ายนัก ปกติความคลาดเคลื่อนจากการสุ่มตัวอย่างจะมีขนาดประมาณ 2-10% ของยอด สถิติที่ต้องการประมาณ หรือบางครั้งอาจจะสูงกว่านั้น ทั้งนี้ขึ้นอยู่กับขนาดของตัวอย่าง โดยทั่วไปความคลาดเคลื่อนจะใหญ่ยิ่งขึ้นเมื่อตารางสถิติ มีการจำแนกรายละเอียดเพิ่มขึ้น

หรือเมื่อต้องการแสดงข้อมูลสถิติเป็นรายเขตปกครอง หรือเขตภูมิศาสตร์ย่อยลงไป เช่น การประมาณยอดข้อมูลเป็นรายอำเภอ หรือรายจังหวัด อาจมีความคลาดเคลื่อนสูงจนไม่สามารถใช้สถิติจากการประมาณให้เป็นประโยชน์ได้ กรณีเช่นนี้เราคงมีแต่เพียงสถิติซึ่งเป็นยอดรวมของประเทศ ซึ่งอาจจะนำมาใช้ไม่ได้ประโยชน์อย่างเต็มที่นัก ข้อนี้เป็นข้อเสียประการสำคัญของวิธีการสำรวจด้วยตัวอย่าง เมื่อเทียบกับวิธีสำมะโนและเมื่อต้องการยอดข้อมูลสถิติเป็นรายเขตปกครองย่อยจริง ก็จะใช้วิธีของการสำรวจด้วยตัวอย่างไม่ได้ หรือถ้าใช้ก็อาจจะต้องเพิ่มขนาดตัวอย่างให้มากเพื่อลดความคลาดเคลื่อนลง

ในเรื่องคุณภาพของข้อมูล เมื่อการสำรวจมีปริมาณการทำงานน้อยกว่าสำมะโน ก็ย่อมหมายความว่าความถูกต้องในการทำงานจะทำได้ดีกว่า แม้แต่พนักงานอาจจะคัดเลือกผู้ที่มีคุณภาพดีกว่าเพราะจำนวนพนักงานที่ใช้มีไม่มากนัก หรือการฝึกอบรมพนักงานเพื่อการทำงานก็จะทำได้อย่างละเอียดถี่ถ้วนดีกว่าที่จะต้องฝึกอบรมพนักงานเป็นจำนวนมาก เพื่อการสำมะโน เฉพาะในส่วนนี้ที่เกี่ยวกับคุณภาพของพนักงานอาจจะช่วยให้ข้อมูลที่เก็บมาได้มีคุณภาพดีกว่าข้อมูลจากสำมะโน หรือในบางครั้งเนื่องจากใช้งบประมาณไม่มากนัก การสำรวจอาจจะใช้เทคนิคการเก็บข้อมูลเบื้องต้นจากสนาม ที่ละเอียดกว่าหรือนิสิตารกว่าวิธีการสำมะโน ซึ่งส่วนนี้ก็จะช่วยให้ข้อมูลมีคุณภาพดีขึ้นกว่าการทำสำมะโน และยังสามารถช่วยรวมข้อมูลบางรายการเพิ่มขึ้น ซึ่งอาจจะไม่สามารถรวบรวมได้โดยวิธีสำมะโน

อีกข้อหนึ่งการสำรวจที่ดีกว่าสำมะโนก็คือ เรื่องคาบเวลาในการทำงาน ข้อมูลจากการสำรวจจะได้มาโดยเร็ว เนื่องจากงานประมวลยอดข้อมูลหลังจากที่เก็บข้อมูลเบื้องต้นจากสนามแล้วมีปริมาณงานไม่มากนัก ข้อมูลที่รวบรวมได้จะสามารถนำมาใช้ได้ทันที ภายในเวลาไม่กี่เดือนหลังจากการปฏิบัติงานสนามแล้ว ซึ่งตรงกันข้ามกับงานสำมะโน ซึ่งจะต้องคอยเป็นเวลา 2-3 ปี กว่าที่จะประมวลข้อมูลแล้วเสร็จ

ในทางปฏิบัติแล้ว งานสำรวจด้วยตัวอย่างจะนำมาใช้เสริมงานสำมะโน เพื่อเก็บข้อมูลสถิติในปีที่ไม่ได้ทำสำมะโน เช่น สำมะโนเกษตร ซึ่งทำในปี 2506 และ 2516 ในระหว่างปีที่ไม่ได้ทำสำมะโน ในขณะที่รัฐมีความจำเป็นต้องใช้ข้อมูลสถิติผลิตผลของพืชสำคัญ เช่น ข้าว ข้าวโพด แต่ละปีสำนักงานสถิติแห่งชาติจะทำการสำรวจผลิตผลเกษตรด้วยตัวอย่าง 20,000 ผู้ประกอบการเกษตร จากจำนวนทั้งหมดทั่วประเทศประมาณ 4 ล้านผู้ประกอบการ หรือกรณีที่ต้องการรวบรวมข้อมูลที่มีรายละเอียดปลีกย่อยเพิ่มเติม ซึ่งไม่สามารถรวบรวมจากสำมะโน

นอกจากนี้งานวิจัยสนาม (field research) บางประเภทก็เป็นการสำรวจด้วยตัวอย่างด้วย งานวิจัยประเภทนี้จะมีการออกไปเก็บข้อมูลเบื้องต้นที่สลับซับซ้อนด้วยการสัมภาษณ์ผู้ตอบ การประมวลผลข้อมูลและการวิเคราะห์ข้อมูล จะทำในเชิงวิจัยเพื่อค้นหาข้อ-ความจริงที่ลึกซึ้งลงไป บางครั้งก็สำรวจเพื่อต้องการข้อมูลสถิติมาใช้อย่างรวดเร็ว เช่น สำรวจในการวิจัยตลาด เป็นต้น

วิธีการรายงานจากหน่วยบริหาร นับว่าเป็นวิธีการหนึ่งที่จะรวบรวมข้อมูลสถิติโดยไม่ต้องสิ้นเปลืองค่าใช้จ่ายในการดำเนินงานทางสถิติมากนัก วิธีการเช่นนี้ใช้กันมาทั้งในสถานประกอบการธุรกิจใหญ่ ๆ และในระหว่างหน่วยงานของรัฐ ในสถานประกอบการธุรกิจก็มี อาทิ เช่น สถิติเกี่ยวกับการผลิต การใช้ทรัพยากรเพื่อการผลิต สถิติแสดงปริมาณการขาย เป็นต้น สถิติเหล่านี้ส่วนมากรวบรวมจากรายงานของหน่วยบริหาร ในระดับชาติ รายงานจากการบริหารเป็นแหล่งผลิตข้อมูลสถิติสำคัญหลายรายการ มี อาทิ เช่น ข้อมูลเกี่ยวกับผลผลิตข้าวและผลผลิตทางการเกษตรอื่น ๆ เช่น รายงานจากผู้ใหญ่บ้าน ผ่านอำเภอ เข้าสู่กรมในกระทรวงเกษตร สถิติภาษีอากรที่จะเก็บได้ซึ่งรายงานโดยสรรพากรจังหวัด เป็นต้น

ข้อมูลที่รายงานจากหน่วยบริหาร มีลักษณะเป็นข้อมูลทุติยภูมิกลาย ๆ คือจะต้องมีการรวบรวมชั้นหนึ่งก่อน (เช่น ในกรณีสถิติการเกษตร ผู้ใหญ่บ้านต้องรวบรวมข้อมูลภายในหมู่บ้านของตนเองก่อน) แล้วจึงส่งมารวมยอดเป็นสถิติระดับชาติ หน่วยที่รายงานข้อมูลมีน้อยหน่วย เมื่อเทียบกับจำนวนหน่วยแฉ่งนับในสำมะโน และการดำเนินงานชั้นรายงานนี้มักไม่ต้องเสียงบประมาณ เพราะผู้รายงานต้องทำหน้าที่ในหน่วยงานและมีเงินเดือนจากงานประจำอื่นอยู่แล้ว การรวบรวมสถิติโดยวิธีนี้จึงใช้งบประมาณเพิ่มเติมแต่เพียงเล็กน้อยเพื่อเป็นค่าใช้จ่ายในการประมวลยอดข้อมูลที่ต้องการ และเพื่อพิมพ์สถิติเผยแพร่เท่านั้น เช่น สถิติเกี่ยวกับการเกษตร หากจะเก็บโดยวิธีสำมะโนจะใช้งบประมาณประมาณ 50-60 ล้านบาท โดยการสำรวจผลผลิตเกษตรประมาณ 3-5 ล้านบาท และโดยรายงานของผู้ใหญ่บ้านจะใช้งบประมาณประมาณ 5 แสนบาท เท่านั้น

ข้อมูลสถิติที่รวบรวมได้โดยวิธีนี้ มักจะมีรายการที่จำกัด เพราะจะต้องให้สอดคล้องกับข้อมูลในแบบของรายงานประการหนึ่ง และเพราะผู้รายงานไม่สามารถ (หรือไม่มีหน้าที่โดยตรง) ที่จะให้ข้อมูลเบื้องต้นในรายละเอียดมากนัก บางครั้งข้อมูลที่รายงานมาเป็นเพียงผลพลอยได้จากการบริหารเท่านั้น ความหวังที่จะได้ข้อรายการสถิติที่มีรายละเอียด-

เสียตมากโดยวิธีรายงานมักจะทำได้

ข้อมูลสถิติที่รวบรวมได้โดยวิธีนี้ มีคุณภาพดีบ้าง ไม่ดีบ้างคละกันไป ทั้งนี้แล้วแต่คุณภาพของข้อมูลเบื้องต้นในแบบรายงาน ข้อมูลเกี่ยวกับปริมาณและราคาสินค้าในใบส่งสินค้าเข้าประเทศ นับว่าเป็นข้อมูลเบื้องต้นที่มีคุณภาพถูกต้องดีพอสมควร (นอกจากน้อยรายที่ทุจริต ที่ใบนำส่งสินค้ากับสินค้าที่นำเข้าจริงไม่ตรงกัน) สถิติประมาณสินค้าเข้าและออกที่ประมวลจากใบนำส่งสินค้าจึงนับว่าเป็นข้อมูลสถิติที่มีคุณภาพดี

แต่รายงานผลผลิตทางการเกษตรจากผู้ใหญบ้านมีคุณภาพไม่ดีนัก เพราะผู้ใหญบ้านอาจจะไม่สามารถรวบรวมยอดสถิติในหมู่บ้านของตน (ก่อนที่จะกรอกแบบรายงาน) ให้ถูกต้อง ในบางกรณีผู้ใหญบ้านไม่เอาใจใส่กลับกรอกตัวเลขที่เตาเอาเอง บางครั้งก็ไม่รายงานตามคำสั่งเสียเฉย ๆ ทำให้เกิดความผิดพลาดในข้อมูลเพิ่มขึ้นอีกจากการรายงานไม่ครบถ้วน และบางครั้งในการคอยรายงานที่ยังเข้ามาไม่ครบยังจะทำให้เกิดความล่าช้าในการประมวลยอดสถิติ ที่ต้องการอีกด้วย

สำหรับข้อมูลสถิติอันเป็นผลพลอยได้จากการทะเบียนนั้น มีลักษณะคล้ายคลึงกับข้อมูลสถิติที่รวบรวมจากรายงาน ผิดกันตรงที่แหล่งเบื้องต้นของข้อมูลซึ่งเป็นเอกสารการทะเบียน ซึ่งการเก็บมีลักษณะเป็นการต่อเนื่อง มีการปรับแก้ให้ถูกต้องทันสมัย สำหรับการเก็บสถิติจากรายงานนั้น เป็นผลมาจากการบริหาร ส่วนมากใช้เพียงครั้งเดียว การใช้ไม่มีลักษณะต่อเนื่อง ข้อมูลสถิติที่เก็บโดยการทะเบียนนั้นเก็บข้อมูลการจำกัดเช่นเดียวกับกรารายงาน ทั้งนี้ข้อมูลสถิติต้องเกิดจากข้อมูลที่เก็บในระบบทะเบียนเท่านั้น จะเกินเลยออกไปไม่ได้ และเท่าที่เป็นอยู่ระบบทะเบียนเป็นระบบข้อมูลที่ค่อนข้างใหญ่ มี พ.ร.บ.คุ้มครอง ทำให้มีความเฉื่อยต่อการปรับเปลี่ยนมาก การที่จะเปลี่ยนระบบทะเบียนเพื่อให้ได้สถิติที่ต้องการย่อมไม่อยู่ในวิสัยที่จะทำได้โดยง่ายนัก คุณภาพของข้อมูลสถิติขึ้นกับคุณภาพของข้อมูลเบื้องต้นที่เก็บในระบบทะเบียน ซึ่งส่วนมากข้อมูลเบื้องต้นเหล่านี้ มักจะไม่ถูกต้องทันสมัยตามความเป็นจริง ทั้งนี้เพราะกลไกที่จะปรับแก้ข้อมูลในทะเบียนให้ทันสมัยมักจะบกพร่องอยู่เสมอ บางครั้งก็เพียงแต่ล่าช้าในการปรับแก้ แต่บางครั้งเจ้าหน้าที่ละเลยปล่อยให้ข้อมูลบางรายการผิดอยู่เรื่อย ๆ โดยไม่ปรับแก้ และนับวันความผิดพลาดก็จะเพิ่มขึ้น

ตัวอย่างข้อมูลสถิติที่จะรวบรวมได้จากระบบทะเบียนราษฎรได้แก่จำนวนประชากรจำแนกเพศ เป็นรายจังหวัด อำเภอและตำบลในวันสิ้นปี ซึ่งประมวลได้จาก (จำนวนประชากรจากทะเบียนเมื่อต้นปี + จำนวนเกิดในระหว่างปี - จำนวนตายในระหว่างปี +

จำนวนคนย้ายเข้าในระหว่างปี - จำนวนคนย้ายออกในระหว่างปี)

เป็นที่เชื่อกันว่าจำนวนประชากรที่รวบรวมได้โดยวิธีนี้เคยไปกว่าความเป็นจริงประมาณร้อยละ 3 - 5 จากการศึกษาตรวจสอบจำนวนประชากรที่อาศัยอยู่จริงในบางหมู่บ้านกับจำนวนที่นับได้ในทะเบียนราษฎรยืนยันข้อเท็จจริงนี้ และการศึกษาายังได้แสดงให้เห็นองค์แห่งความผิดพลาดของทะเบียนยิ่งไปกว่านั้นอีก ถึงแม้ว่ายอดสถิติจากทะเบียนจะน้อยกว่าเป็นจริงเพียงร้อยละ 3-5 แต่บุคคลที่อาศัยอยู่จริง ยืนยันได้กับชื่อในทะเบียนได้เพียงร้อยละ 75 เท่านั้น อีกร้อยละ 20 มีรายชื่ออยู่ในทะเบียนบ้าน แต่ตัวจริงได้ย้ายออกไปจากท้องที่เสียแล้ว และขณะเดียวกันก็มีอีกส่วนหนึ่งอีกร้อยละ 25 มีตัวตนอยู่ในท้องที่แต่ยังไม่ได้อายทะเบียนเข้ามา ส่วนที่เกินทะเบียนกับส่วนที่ขาดทะเบียนหักลบกันไปทำให้ยอดสถิติขาดไปเพียงร้อยละ 5 ก็ยังนับว่าเป็นสถิติที่พอจะนำมาใช้เป็นประโยชน์ได้บ้าง

สำหรับจำนวนการเกิดและการตายก็เคยไปกว่าความเป็นจริง เช่น จากการแจ้งการเกิดในระบบทะเบียน อัตราการเกิดของประชากรเท่ากับร้อยละ 3.6 แต่จากการสำรวจด้วยตัวอย่าง² อัตราเกิดเท่ากับร้อยละ 4.3 คือทะเบียนบันทึกการเกิดน้อยไปเกือบร้อยละ 20 ในขณะที่การบันทึกการตายจากทะเบียนได้อัตราการตายเท่ากับร้อยละ 0.8 แต่จากการสำรวจด้วยตัวอย่างได้อัตราการตายร้อยละ 1.2 ทะเบียนบันทึกการตายน้อยไปกว่าร้อยละ 30 ความผิดพลาดในการบันทึกการเกิดและการตายหักลบกันไปส่วนหนึ่งในตอนคำนวณจำนวนประชากรที่เพิ่ม แต่ถึงกระนั้นอัตราการเพิ่มของประชากรเท่ากับ (อัตราการเกิด - อัตราการตาย) คำนวณจากการทะเบียนก็ยังไม่ต่ำกว่าที่สำรวจได้ คือจากทะเบียนได้อัตราเพิ่มเท่ากับร้อยละ 2.8 แต่จากการสำรวจได้ร้อยละ 3.1 สำหรับเรื่องการย้ายเข้า/ออกยังไม่เคยได้มีการศึกษากันอย่างจริงจัง แต่เข้าใจว่าการแจ้งย้ายเข้าหรือออก (ซึ่งประชาชนมีหน้าที่ต้องทำตามกฎหมาย) คงไม่ได้ทำอย่างครบถ้วน และตรงตามเวลาตามความต้องการของนักกฎหมายนัก ทั้งนี้ได้แปลความหมายว่าข้อมูลสถิติรวบรวมจากระบบทะเบียนจะผิดพลาดเช่นนี้เสมอไป หรือเมื่อผิดแล้วจะนำสถิติมาใช้เป็นประโยชน์ไม่ได้ ความผิดพลาดในหลายชั้นกระบวนการอาจจะหักลบกันไปบ้าง ทำให้ยอด

²สำรวจการเปลี่ยนแปลงของประชากร

รวมไม่ผิดไปมากนัก (ดังเช่น สถิติประชากรจากทะเบียนราษฎร) และถึงหากสถิติจะผิดพลาดก็ยังสามารถนำมาใช้เป็นประโยชน์ได้ ระบบทะเบียนบางระบบอาจจะทำให้เกิดข้อมูลสถิติที่มีความแม่นยำพอประมาณ และความจริงเป็นวิธีการเดียวที่จะให้ข้อมูลสถิติในเรื่องนั้น ๆ เช่น การจดทะเบียนยานพาหนะที่กรมตำรวจ มีองค์แห่งความแม่นยำสูง ทำให้ประมวลสถิติจำนวนยานพาหนะประเภทต่าง ๆ ได้อย่างถูกต้อง

สรุปผลดี/ผลเสียของวิธีการรวบรวมสถิติทั้ง 4 วิธีดังนี้

<u>วิธีสำมะโน</u>	<u>ข้อดี</u>	<u>ข้อเสีย</u>
	1) ได้ข้อมูลครบถ้วน แสดงรายละเอียด ข้อมูลเป็นเชิงภูมิศาสตร์ที่เล็กที่สุดได้	1) สิ้นเปลืองค่าใช้จ่ายมาก
	2) เป็นข้อมูลหลัก ซึ่งจะนำไปช่วยในการวางแผน การเก็บข้อมูลอื่น ๆ	2) ไม่สามารถทำได้ทุกปี ทำให้ข้อมูลไม่ต่อเนื่องกัน
		3) ปริมาณงานมาก ทำให้ยากแก่การบริหารงานคุณภาพในข้อมูลอาจจะหย่อนลงไปเพราะเหตุนี้
		4) ล่าช้าในการประมวลผล

การสำรวจ

1) สิ้นเปลืองค่าใช้จ่ายน้อย	1) ข้อมูลเป็นข้อมูลที่ประมาณชั้น มีความคลาดเคลื่อนในข้อมูลจากการสุ่มตัวอย่าง
2) สามารถทำได้บ่อย ได้ข้อมูลที่ต่อเนื่อง ใช้สนับสนุนการสำมะโนในปีที่ไม่มีการทำสำมะโนได้	2) ไม่สามารถประมาณข้อมูลเป็นรายเชิงภูมิศาสตร์ที่เล็กได้หรือไม่สามารถจำแนกข้อมูล

ข้อดี

- 3) ปริมาณงานน้อยจึงควบคุมได้ใกล้ชิด
- 4) รวบรวมข้อมูลที่สลับซับซ้อนหรือพิสดารได้ เพราะอาจใช้เทคนิคการวัดหรือสัมภาษณ์พิเศษได้
- 5) ประมวลข้อมูลสำเร็จรูปได้อย่างรวดเร็ว ทันต่อการใช้

ข้อเสีย

ในรายละเอียดมากนัก
เพราะมีความคลาดเคลื่อนมาก

วิธีรายงาน

- | | |
|--|--|
| <ol style="list-style-type: none"> 1) สิ้นเปลืองค่าใช้จ่ายน้อย เพราะเสียค่าใช้จ่ายเฉพาะส่วนประมวลข้อมูลกับการพิมพ์เผยแพร่เท่านั้น | <ol style="list-style-type: none"> 1) รายการของข้อมูลจำกัด เพราะต้องรวบรวมขึ้นตามแบบรายงาน ซึ่ง เป็นผลพลอยได้จากการบริหาร |
|--|--|

ในบางครั้งหากมีความจำเป็นต้องใช้ข้อมูลตามความเป็นจริง ในขณะใดขณะหนึ่ง ก็ใช้เวลานั้น ๆ เป็นเวลาที่ข้อมูลอ้างถึงได้ เช่น การสำรวจเพื่อวิจัยตลาดเมื่อฝ่ายขายได้ทำการโฆษณาอย่างเข้มข้นในขณะใดขณะหนึ่ง อาจจะต้องการประเมินผลการโฆษณาด้วยการวิจัยตลาด ซึ่งในที่นี้ก็จะต้องทำทันทีภายหลังจากการโฆษณานั้น ๆ

การเตรียมงบประมาณ

โครงการเก็บรวบรวมข้อมูลสถิติ ไม่ว่าจะ เป็นด้วยวิธีใด เป็นงานที่จะเสียค่าใช้จ่ายเป็นจำนวนมาก ซึ่งเมื่อมองอย่างผิวเผิน ดูเหมือนว่าจะเป็นการใช้จ่ายโดยเปล่าประโยชน์ เพราะไม่มีผลงานประจักษ์เป็นประโยชน์อย่างเด่นชัด และเป็นการใช้จ่ายที่ไม่มีความจำเป็นอันเร่งด่วน ไม่ว่าจะ เป็นระดับชาติ หรือระดับหน่วยงาน เมื่อถึงเวลาที่มี

ความจำเป็นจะต้องตัดรายจ่าย ส่วนมากมักจะมุ่งการตัดมาที่โครงการเก็บรวบรวมสถิติเสมอ

หมายความว่านักสถิติ จะต้องใช้ความพยายามเป็นพิเศษที่จะอธิบายถึงความจำเป็นของการที่จะต้องมีข้อมูล และประโยชน์ที่จะได้รับจากการใช้ข้อมูลเหล่านั้น

ผ่านขั้นนี้ไปได้แล้ว นักสถิติจะต้องกำหนดงบประมาณในรายละเอียดในการที่จะทำงานให้เสร็จสิ้นทั้งโครงการ ในทางปฏิบัติการเตรียมงบประมาณจะเป็น 2 ชั้น ๑ แรกเป็นงบประมาณของการดำเนินตามแผนที่นักสถิติวางไว้ เมื่อเสนองบประมาณไป มักจะถูกตัดออกไปส่วนหนึ่ง ก็เป็นหน้าที่ของนักสถิติที่จะต้องกลับมาปรับวิธีดำเนินงาน หรือปรับปริมาณงานจนกระทั่งจะสามารถทำงานให้เสร็จสิ้นได้ในวงเงินที่กำหนดมาให้ เช่นในการสำรวจอาจจะต้องลดขนาดของตัวอย่างลง โดยยอมให้ความคลาดเคลื่อนในข้อมูลสูงขึ้น งบประมาณที่จะใช้ในงานสำมะโน/สำรวจหนึ่ง ๆ อาจจะแบ่งออกได้เป็นขนาดใหญ่ ๆ คือ

(1) ขณะเตรียมงาน ส่วนมากเป็นค่าจ้างเจ้าหน้าที่วิชาการ พนักงานเพื่อทดสอบการทำงาน อาจจะมีค่าเดินทาง ค่าวัสดุ อุปกรณ์งบประมาณ

(2) ค่ากระดาษ ค่าพิมพ์ใบแบบสอบถาม และข้อสั่งชี้แจง ซึ่งจะต้องทราบก่อนล่วงหน้าว่าจะใช้ใบแบบสอบถามกี่แผ่น (ปกติต้องพิมพ์เกินไว้ประมาณร้อยละ 15) ข้อสั่งชี้แจงที่เล่ม

(3) ค่าใช้จ่ายในการแจกจ่ายของพนักงานแจกจ่าย ซึ่งส่วนใหญ่เป็นค่าจ้าง ซึ่งอาจจะเป็นการจ้างเป็นรายชิ้นของผลงาน เช่น รายบ้านที่นับจดมาได้ = บ้านละ 2 บาท หรือรายบ้านที่แจกจ่ายมาได้ = บ้านละ 12 บาท เป็นต้น หรืออาจจะเป็นการจ้างเหมาทั้งโครงการ หรืออาจจะเป็นการจ้างเป็นรายวันก็ได้ นอกจากนี้ ก็มีส่วนที่จะใช้จ่ายเป็นค่าพาหนะในการเดินทางไปยังศูนย์ เพื่อรับการอบรม ค่าเดินทางไปปฏิบัติงานในท้องที่ ค่าพาหนะเองก็ยังสามารถจ่ายได้หลายแบบ เช่น จ่ายเหมาเป็นรายวัน จ่ายให้ตามที่พนักงานใช้ไปจริงก็ได้ ในบางท้องที่ที่ทำงานค่าพาหนะอาจจะสูงมาก และการที่จะควบคุมการใช้จ่ายรายการนี้ ให้เป็นไปอย่างมีประสิทธิภาพจะทำได้โดยไม่ง่ายนัก นอกจากนี้จะต้องเตรียมงบประมาณเพื่อจ่ายให้พนักงานเป็นค่าที่พัก ค่าเบี้ยเลี้ยง ฯลฯ บางครั้งอาจจะต้องตั้งงบประมาณเป็นค่าสมนาคุณแก่ผู้มาช่วยราชการ เช่น นายอำเภอ ช่วยควบคุมการปฏิบัติงานสนาม

ผู้วางแผนการปฏิบัติงาน จะต้องทดสอบวิธีดำเนินการในแบบต่าง ๆ อาจจะใช้วิธีการจ่ายค่าตอบแทนต่าง ๆ กัน ในอัตราที่ต่างกัน และต้องประเมินประสิทธิภาพของการทำงาน ประกอบกับคุณภาพของผลงานที่ได้ และจะเลือกแบบหนึ่งที่ใช้จ่ายน้อยที่สุด และให้ผลงานที่มีคุณภาพในระดับที่ยอมรับเป็นที่เชื่อถือได้

(4) ค่าใช้จ่ายในการประมวลข้อมูล ซึ่งจะต้องใช้เป็นค่าจ้างนักวิชาการและพนักงานส่วนหนึ่ง และวัสดุครุภัณฑ์อีกส่วนหนึ่ง เช่นค่ากระดาษพิมพ์รายงาน ค่าเช่าเครื่องจักรคำนวณ ค่าเครื่องพิมพ์ดีด โตะ แก้ว ค่าเครื่องคิดเลข ฯลฯ ในการทำงานขั้นนี้นักสถิติก็ต้องพยายามเลือกแบบการทำงานที่สามารถให้ได้ผลงานที่ต้องการ โดยใช้งบประมาณน้อยที่สุด

(5) ค่าพิมพ์รายงานสถิติเพื่อเผยแพร่ ซึ่งจะต้องประมาณไว้ก่อนว่าจะมีกี่รายงาน กี่ตารางสถิติ กี่หน้าและจะพิมพ์รายงานละกี่เล่ม

ในงานสำมะโน/สำรวจหนึ่ง ๆ ค่าใช้จ่ายในการเก็บข้อมูลเบื้องต้นในสนามจะตกประมาณ 50 เปอร์เซ็นต์ของรายจ่ายทั้งหมด 25 เปอร์เซ็นต์ เป็นค่าใช้จ่ายในการประมวลผล ส่วนอีก 25 เปอร์เซ็นต์ที่เหลือ เป็นค่าใช้จ่ายในการพิมพ์แบบสอบถามแบบฟอร์ม และพิมพ์รายงานเพื่อเผยแพร่ และค่าจ้างเจ้าหน้าที่วิชาการอื่น ๆ

การวางแผนและดำเนินการสำรวจข้อมูลสถิติโดยวิธีสำรวจตัวอย่างขนาดใหญ่

ดร.นิยม ปุราคำ อดีตเลขาธิการสำนักงานสถิติแห่งชาติ ได้กำหนดขั้นตอนที่สำคัญ ๆ ไว้ทั้งหมด 13 ขั้นตอน ในการวางแผนและดำเนินการสำรวจข้อมูลสถิติโดยระเบียบวิธีตัวอย่างขนาดใหญ่ (large-scale sample survey) ดังนี้

(1) กำหนดวัตถุประสงค์ (objective) นับเป็นขั้นแรกของงานที่จะต้องจัดทำคือการระบุวัตถุประสงค์หรือเป้าหมายที่แน่นอนชัดเจนของโครงการสำรวจว่ามีวัตถุประสงค์อย่างไร? จากประสบการณ์ของนักสถิติอาชีพส่วนใหญ่ซึ่งเกี่ยวข้องกับการให้คำแนะนำทางวิชาการสถิติ พบว่ามีหน่วยงานหลายหน่วยงานที่มีความประสงค์จะทำการสำรวจข้อมูลสถิติ และมาขอคำแนะนำปรึกษาทางเทคนิคจากนักสถิติ ไม่สามารถบอกวัตถุประสงค์หรือความต้องการที่แน่นอนของการสำรวจและบางกรณียังไม่ทราบด้วยซ้ำว่าจะนำผลการสำรวจ (survey results) ไปใช้ทำอะไรอย่างไร? ถ้าไม่อาจจะระบุวัตถุประสงค์ได้แน่นอนชัดเจนแล้ว ก็เป็นที่น่าสงสัยว่าคุณค่าของการสำรวจจะมีไม่มากนัก เพราะ

นักสถิติก็มักจะไม่สามารถให้คำแนะนำทางด้านวิชาการได้อย่างถูกต้อง และอาจจะวางแผนการสำรวจ (survey design) ที่ให้ผลที่ไม่ตรงกับวัตถุประสงค์หรือความต้องการทำให้เสียเงินเสียกำลังงานและเวลาของทุกฝ่ายไปเปล่า ๆ ฉะนั้นในการวางแผนการสำรวจข้อมูลสถิติใด ๆ ก็ตามจะต้องเริ่มต้นด้วยการระบุวัตถุประสงค์ที่แน่นอนชัดเจน เช่น

ชื่อโครงการ การสำรวจต้นทุนการผลิตข้าวของชาวนา

วัตถุประสงค์ (1) เพื่อรวบรวมข้อมูลเกี่ยวกับค่าใช้จ่ายที่ประกอบกันเป็นต้นทุนการผลิตข้าวต่อไร่ และต่อเกวียน เพื่อการประมาณต้นทุนรวม (total cost) และองค์ประกอบที่สำคัญของต้นทุน (major cost components) ของการผลิตข้าวในภาคต่างๆ ในแต่ละปี รวมทั้งศึกษาแนวโน้มของต้นทุนซึ่งจะเป็นประโยชน์ต่อการกำหนดนโยบายประกันราคาขั้นต่ำ (floor price)

(2) ระบุประชากรที่อยู่ในข่ายการสำรวจ (population to be covered) อาจนับเป็นส่วนหนึ่งของการกำหนดวัตถุประสงค์ของการสำรวจ

คือนอกจากจะต้องระบุเป้าหมายที่แน่นอนดังที่กล่าวในข้อ (1) แล้ว การวางแผนการสำรวจจะต้องระบุประเภทของประชากรที่จะครอบคลุม (cover) ในการสำรวจด้วย เช่น การสำรวจสถิติการเกษตร (Agricultural Survey) เราอาจจะพยายามแจงนับ (enumerate) ครัวเรือนที่ประกอบการเกษตรทุกขนาด ไม่ว่าจะมีส่วนที่เพาะปลูกหรือกิจกรรมทางเกษตรมากน้อยเพียงไร หรืออาจจะจำกัดการสำรวจเฉพาะครัวเรือนเกษตรที่มีเนื้อที่การเกษตรไม่น้อยกว่า 2 ไร่ หรือมีรายได้จากการขายผลิตผลทางการเกษตรไม่ต่ำกว่าปีละ 1,800 บาท เป็นต้น เช่นเดียวกับการสำรวจอุตสาหกรรม (Industrial Survey) อาจจะไม่จำเป็นต้องแจงนับสถานการณ่อุตสาหกรรมหรือโรงงานขนาดเล็กๆ ซึ่งมีคนงานไม่เกิน 2 คน หรือการสำรวจรายได้-รายจ่ายของครัวเรือน อาจจะไม่รวมครัวเรือนสถาบัน (institution households) เช่น หอพัก โรงพยาบาล โรงแรม ฯลฯ เป็นต้น การระบุประชากรที่อยู่ในข่ายที่จะเกี่ยวข้องกับการเลือกตัวอย่าง (sample selection) เพื่อให้ได้ตัวแทนของประชากรที่ต้องการด้วย ถ้าผู้ที่จะดำเนินการสำรวจยังไม่สามารถระบุประชากรที่อยู่ในข่ายได้แน่นอน ผู้วางแผนการสำรวจ (survey designer) ก็คงจะไม่สามารถเลือกตัวอย่าง (select sample) ที่จะเป็นตัวแทนของประชากรที่ต้องการได้

(3) การเตรียม sampling frame : เมื่อได้ระบุประชากรที่อยู่ในข่ายที่จะทำการสำรวจแล้ว ขั้นตอนไปก็คือการหารายชื่อ (list) หรือทะเบียน หรือทำเนียบ (ถ้ามี) ของหน่วยทุก ๆ หน่วยที่ประกอบกันเป็นประชากรที่ต้องการศึกษา ซึ่งเรียกกันว่า sampling frame ในทางปฏิบัติสิ่งที่อาจใช้เป็น sampling frame อาจเป็นสิ่งที่มีอยู่แล้วในหน่วยงานต่างๆ เช่น รายชื่อโรงงาน รายชื่อหมู่บ้าน ทำเนียบสถาบันกลุ่ม ชาวนา หรือแผนที่แสดงอาณาเขตที่จะสามารถระบุตำแหน่งในระดับต่างๆ ได้ หรืออาจจะต้องทำขึ้นมาใหม่โดยการสำรวจพิเศษ (special survey) เช่น ส่งแบบให้ผู้ใหญ่บ้าน ช่วยกันกรอกชื่อและที่อยู่ของ โรงงานที่ต้องการสำรวจในหมู่บ้าน เข้ามาก่อนการสำรวจจริง 3 เดือน ฯลฯ ในการเตรียมหรือจัดทำ frame สำหรับการสำรวจนั้น ควรคำนึงให้เป็น frame ที่มีลักษณะต่อไปนี้

ก. เป็น frame ที่สมบูรณ์ไม่ขาดไม่เกิน เช่นถ้าใช้สมุดโทรศัพท์เป็น frame ในการสำรวจเกี่ยวกับรายได้รายจ่ายของประชากรในพระนคร-ธนบุรี ก็จะไม่ใช้ frame ที่สมบูรณ์ เพราะมีครัวเรือนจำนวนกว่าครึ่งที่ไม่มีรายชื่ออยู่ในสมุดโทรศัพท์ ยิ่งกว่านั้นบางครัวเรือนกลับมีชื่อปรากฏมากกว่า 1 ครั้ง เพราะมีหลายบ้านหลายเบอร์

ข. เป็น frame ที่ทันสมัยหรือจัดทำขึ้นล่าสุด เช่น ไม่ใช่ไปเอาทะเบียนหรือทำเนียบโรงสีหรือโรงงานอื่นๆ ซึ่งปรับปรุงล่าสุดเมื่อปี 2506 มาใช้เป็น frame ในการสำรวจที่จะทำในปี 2514 เพราะถ้าทำเช่นนั้นก็คงจะปรากฏว่ามีโรงงานหลายโรงที่ปรากฏในทะเบียนหรือทำเนียบที่เป็น frame แต่เมื่อส่งพนักงานออกไปสำรวจปรากฏว่าไม่มีตัวตน เพราะเลิกกิจการหรือหยุดไปแล้วเป็นจำนวนมาก และในขณะเดียวกันก็คงจะมีโรงงานใหม่ ๆ เกิดขึ้นอีกไม่น้อยที่ไม่อยู่ในข่ายของการสำรวจ เพราะไม่ได้ปรากฏใน frame ทำให้การสำรวจที่ทำในปี 2514 ได้ข้อมูลที่ไม่ใช่ข้อเท็จจริงซึ่งอาจผิดวัตถุประสงค์ของการสำรวจไปอย่างมากและอาจเป็นผลร้ายทำให้ผู้ใช้ข้อมูล (users of data) ได้ข้อสรุปที่ผิดพลาดจนกระทั่งตัดสินใจหรือวางนโยบายที่ผิดพลาดตามไปด้วย ซึ่งนับเป็นผลงานที่ฝรั่งเขาเรียกว่า worse than useless กล่าวคือเลวเสียยิ่งกว่าไม่มีประโยชน์เพราะทำให้เกิด misleading ต่าง ๆ นานา

ค. เป็น frame ที่มีข้อมูลที่จะบอกตำแหน่งที่หรือที่อยู่ของหน่วยนั้น ทั้งนี้เพราะในขั้นของการปฏิบัติงานสนาม (field enumeration) ไม่ว่าจะ เป็นโดยวิธีการส่งพนักงานออกไปสัมภาษณ์ (interviewing method) หรือส่งแบบสอบถามไปให้

ผู้ตอบกรอกข้อมูลเองทางไปรษณีย์ (mail questionnaire method) จะต้องมี address ที่สามารถหาหรือติดต่อกับหน่วยนั้น ๆ ได้ มิฉะนั้น frame ที่มีซึ่งแม้จะสมบูรณ์ด้วยคุณลักษณะในข้อ ก. และ ข. ก็อาจจะไม่มีประโยชน์ในการใช้แต่ประการใด

อนึ่ง เพื่อให้ได้มาซึ่ง frame ที่สมบูรณ์ด้วยคุณลักษณะในทั้ง 3 ข้อ ดังกล่าวแล้วในการทำการสำรวจข้อมูลสถิติเราอาจจะต้องลงทุนปรับปรุง sampling frame ดังกล่าวเสียขั้นหนึ่งก่อน การปรับปรุงนี้อาจใช้วิธีการต่าง ๆ ที่เหมาะสม เช่น ทาหลักฐานที่เกี่ยวข้องกับการเปลี่ยนแปลงของหน่วย (unit) ที่ประกอบกันเป็นประชากร เช่น ทาหลักฐานการจดทะเบียนเพิ่มเติมหรือเลิกกิจการของโรงงาน การเปลี่ยนแปลงอาณาเขตของท้องที่ ฯลฯ และในบางกรณีอาจจะต้องทำการสำรวจสนาม (field survey) เพื่อหาข้อมูลเบื้องต้นเกี่ยวกับ unit ที่ควรจะอยู่ในข่ายของการสำรวจ ซึ่งเรียกกันว่า listing survey หรือ listing ซึ่งเป็นการสำรวจเพื่อบันทึกและบันทึกจำนวนชื่อและที่ตั้งของแต่ละหน่วยและอาจมีรายละเอียดอื่น ๆ อีก 2-3 รายการที่จะใช้ช่วยในการกำหนดตัวอย่างในการสำรวจจริง listing survey นี้ควรทำโดยมีแผนที่ประกอบเพื่อจะให้เห็นใจได้ว่าครอบคลุมท้องที่ทั้งหมด และอาจถือโอกาสระบุตำแหน่งคร่าว ๆ ไว้ในแผนที่เพื่อให้สะดวกในการสำรวจขั้นต่อไป ซึ่งอาจทำโดยพนักงานอีกชุดหนึ่งด้วย

(4) กำหนดหน่วยแจงนับ (enumeration units) ส่วนหนึ่งของ การวางแผนสำรวจข้อมูลสถิติก็คือการกำหนดหน่วยแจงนับ (enumeration units) ซึ่งหมายถึงหน่วยที่เราจะส่งพนักงานไปพบและสัมภาษณ์ เพื่อให้ได้ข้อมูลที่ต้องการ การสำรวจ บางอย่างหน่วยแจงนับอาจเป็นครัวเรือนและการสัมภาษณ์อาจทำกับหัวหน้าครัวเรือนคนเดียวก็อาจได้ข้อมูลทั้งหมด เช่น การสำรวจเกี่ยวกับผลผลิตการเกษตร การสำรวจเกี่ยวกับการใช้แรงงาน การสำรวจรายได้-รายจ่ายของครัวเรือน ฯลฯ แต่การหาข้อมูลในการสำรวจบางอย่างอาจต้องสัมภาษณ์ทุกคนในครัวเรือน เช่น การสำรวจทัศนคติของประชากรและสำรวจปริมาณการใช้สินค้าบางอย่างโดยเฉพาะประเภท consumer products เช่น เครื่องสำอาง ซึ่งแต่ละคนในครัวเรือนเดียวกันอาจใช้ของไม่เหมือนกัน เกี่ยวกับหน่วยแจงนับและผู้ให้คำตอบ (respondent) นี้จะต้องมีแผนงานสำรองไว้ว่าถ้าไม่สามารถจะสัมภาษณ์หัวหน้าครัวเรือนซึ่งเป็น primary respondent ได้ควรจะสัมภาษณ์ใครซึ่งใช้เป็น secondary respondent แทน เช่น ให้สัมภาษณ์บุคคลอื่นในครัวเรือนนั้นซึ่งอายุตั้งแต่ 18 ปีขึ้นไปแทน เป็นต้น

(5) การกำหนดรายการข้อมูลที่จะรวบรวม (information to be collected) นับเป็นสิ่งที่ต้องทำเป็นอันดับแรกอย่างหนึ่งในการวางแผนสำรวจข้อมูลสถิติ เพราะถ้าจำนวนและประเภทของข้อมูลที่จะรวบรวมมีมากและสลับซับซ้อนเกินความจำเป็นแล้ว ในการปฏิบัติงานสนาม (field work) จะประสบปัญหาอย่างมาก เช่น ผู้ตอบ (respondents) อาจจะเกิดความเบื่อหน่ายที่จะต้องเสียเวลาตอบคำถามยาว ๆ จนกระทั่งอาจรีบ ๆ ตอบเพื่อให้การสัมภาษณ์เสร็จ ๆ ไปโดยไม่คำนึงว่าผู้สำรวจจะได้ข้อมูลที่ถูกต้องหรือไม่ หรือยิ่งกว่านั้นอาจมีปฏิกิริยามากจนไม่ยอมตอบคำถามเอาเลย ฉะนั้นในการกำหนดรายการข้อมูลที่จะรวบรวมนี้ควรยึดหลักว่าจะถามเพื่อให้ได้ข้อมูลที่ตรงกับวัตถุประสงค์ของการสำรวจเท่านั้น แต่ก็ไม่ต้องถึงกับจำกัดจำนวนคำถามจนกระทั่งขาดรายการที่สำคัญไป ในกรณีที่ยังไม่ทราบว่าควรจะถามอะไรบ้าง เพื่อให้ได้ข้อมูลครบตามวัตถุประสงค์ของการสำรวจ ก็อาจจะใช้วิธีการร่างหัวข้อ (outlines) ของตาราง (tables) ประมวลผลข้อมูลที่ต้องการจะได้ขึ้นมาก่อนแล้วค่อย ๆ เขียนรายการคำถามที่จะทำให้ได้ข้อมูลในตารางที่ร่างไว้ ซึ่งจะช่วยให้สามารถกำจัดรายการคำถามที่ไม่จำเป็นหรือไม่ตรงออกไปได้มาก และในขณะเดียวกันก็ทำให้แน่ใจได้ว่าจะไม่ตกคำถามที่จำเป็นอันหนึ่งอันใดไป เรื่องสำคัญที่ไม่ควรลืมอีกอย่างหนึ่งก็คือต้องคำนึงถึงวิสัยสมารถของผู้ตอบด้วยว่ารายการข้อมูลที่จะรวบรวมในการสำรวจนี้ ผู้ตอบจะสามารถตอบได้ไม่ใช่ไปพยายามถามสิ่งที่ผู้ตอบไม่มีทางจะให้คำตอบที่แน่นอนได้

(6) กำหนดวิธีการแจกแจงหรือวิธีการเก็บรวบรวมข้อมูล (method of collecting information) เมื่อได้กำหนดรายการข้อมูลที่จะรวบรวมแล้ว ขั้นตอนต่อไปก็คือการกำหนดวิธีการเก็บรวบรวมข้อมูลหรือที่เรียกว่าวิธีการแจกแจง (enumeration method) เช่นการแจกแจงนี้จะใช้วิธีการส่งแบบสอบถามไปให้ผู้ตอบกรอกเองทางไปรษณีย์ (mail interview) หรือจะแจกแจงโดยการส่งเจ้าหน้าที่ไปสัมภาษณ์ด้วยตัวเอง (personal interview) ซึ่งจะเลือกใช้วิธีไหนก็ต้องคำนึงถึงความเหมาะสมทั้งในด้านวิสัยสมารถของผู้ตอบ (respondents) ค่าใช้จ่าย (cost) ในการดำเนินงานและความถูกต้อง (accuracy) ของข้อมูลที่ต้องการ ถ้ากล่าวโดยทั่ว ๆ ไปแล้วอาจถือเป็นหลักปฏิบัติได้ว่าถ้าสามารถทำได้ วิธีการเก็บรวบรวมข้อมูลโดยการส่งเจ้าหน้าที่ไปสัมภาษณ์ผู้ตอบ (personal interviewing method) จะมีประสิทธิภาพกว่าวิธีการส่งแบบไปรษณีย์ให้ผู้ตอบกรอกแล้วส่งกลับคืนมาเอง เพราะพนักงานสำรวจจะได้มีโอกาสสังเกตการณ์ด้วย

ตัวเอง (physical observation) ไปในตัวด้วย และปัญหาการไม่ได้รับคำตอบกลับคืนซึ่งเรียกกันว่า non-response จะมีน้อยกว่า ระเบียบการสำรวจโดย self-enumeration หรือ mailing out questionnaires method ซึ่งนอกจากปัญหาการไม่ได้รับคำตอบแล้วยังอาจมีปัญหาการกรอกข้อมูลไม่สมบูรณ์ หรือไม่ตรงตามความมุ่งหมายของการสำรวจได้ โดยเฉพาะถ้าแบบสอบถามที่ใช้นั้นยากเกินไปและไม่มีคำอธิบายหรือข้อชี้แจง (instruction) ที่เพียงพอในการเก็บรวบรวมข้อมูลบางอย่าง เช่น ข้อมูลเกี่ยวกับผลผลิตเฉลี่ยต่อไร่ (yield) หรือข้อมูลเกี่ยวกับการบริโภคอาหารอาจจะต้องใช้วิธีการวัดแบบวัตถุวิสัย (objective measurement technique) เช่น ตั้งแปลงเก็บเกี่ยวหาผลผลิตกันจริง ๆ หรือชั่งตวงวัดปริมาณอาหารที่บริโภคกันจริง ๆ เพื่อให้ได้ข้อมูลที่ใกล้เคียงกับความเป็นจริงมากที่สุด

(7) กำหนดเวลาอ้างอิง (time reference) หรือคาบเวลาสำรวจ (survey period) ที่แน่นอน เช่น การสำรวจเกี่ยวกับจำนวนปลูสดัตว์ หรือปริมาณเสด็จคผลผลิตผลการเกษตรซึ่งอาจต้องใช้เวลาในการสำรวจหลายวัน เช่น 2 สัปดาห์ ผู้วางแผนการสำรวจจะต้องกำหนด time reference ให้แน่นอนว่าต้องการตัวเลขเสด็จคเมื่อวันที่เท่าไร เพราะอาจมีการเคลื่อนย้ายหรือเปลี่ยนแปลงหลังจากนั้นทำให้เกิดเคลื่อนย้ายหรือเปลี่ยนแปลงหลังจากนั้นทำให้เกิดปัญหานับซ้ำ (double counting) หรือนับขาด (omission) ได้ หรือการสำรวจเกี่ยวกับการบริโภคอาหาร (food consumption survey) จะต้องเลือกพิจารณาว่าจะใช้ time reference เป็น 1 วัน หรือ 1 สัปดาห์ หรือ 1 เดือน ทั้งนี้ก็แล้วแต่ว่าช่วงเวลาไหนจะเป็นตัวแทนที่ดีที่สุดและให้ข้อมูลที่ใกล้เคียงความเป็นจริงที่สุด อย่างไรก็ตามก็มีข้อสังเกตว่า time-reference ที่ใช้ไม่ควรถายาวจนเกินไปนัก เพราะอาจมีปัญหาคความคลาดเคลื่อนเพราะความหลงลืมของผู้ให้ข้อมูล (memory lapse) มาก

(8) การกำหนดและสร้างแบบสอบถามสำหรับการสำรวจ (preparation of the questionnaire or the survey schedule) นับเป็นเรื่องที่มีความสำคัญอย่างหนึ่งในการวางแผนการสำรวจข้อมูลสถิติต่างๆ อยู่มาก เพราะอาจจะทำให้ได้ข้อมูลที่ผิดที่ควร ถ้าแบบสอบถามหรือ questionnaire ที่ใช้ในการสำรวจไม่เหมาะสมไม่ว่าจะเป็นแบบที่สร้างขึ้นเพื่อให้ผู้ที่จะต้องให้ข้อมูลกรอกเองซึ่งเรียกว่า self-enumeration form หรือเป็นแบบที่จะให้พนักงานใช้สัมภาษณ์โดยตรงซึ่งเรียกว่า inter-

viewing form ในทั้งสองกรณีมีหลักสำคัญๆ ที่อาจจะใช้เป็นแนวในการสร้างแบบสอบถาม คือ ควรเริ่มต้นด้วยการเขียนรายการข้อมูลที่ต้องการถามขึ้นก่อนแล้วจึงค่อยๆ เรียบเรียงเป็นคำถามโดยคำนึงให้คำถามนั้นเป็นคำถามที่ชัดเจน (clear) ไม่กำกวม (unambiguous) และตรงประเด็น (to the point) มากที่สุด ฟังระลึกอยู่เสมอว่าคำถามที่คลุมเครือ (vague) จะไม่ทำให้ได้คำตอบที่ชัดเจนไปกว่าตัวคำถามเอง และควรพยายามหลีกเลี่ยงคำถามนำ (leading question) เพราะอาจจะทำให้ได้คำตอบที่มีความเอนเอียง (biased answer) จากผู้ตอบ ข้อที่ควรระวังอีกอย่างหนึ่งในการสร้างแบบสอบถามก็คือเรื่องลำดับ (order) ของคำถาม ควรจะเป็นไปในลักษณะที่ธรรมชาติมากที่สุดหรือถามให้เสร็จไปเป็นเรื่อง ๆ ไม่ใช่ถามเรื่องหนึ่งยังไม่ทันจบก็กลับไปถามเรื่องอื่น แล้ววนกลับมาถามเรื่องเดิมอีกซึ่งอาจทำให้เกิดความสับสนหรือความไม่พอใจแก่ผู้ตอบ แบบสอบถามที่ร่างขึ้นก่อนจะเอาไปใช้จริงควรจะได้มีการทดสอบ (test) เพื่อศึกษาความเหมาะสมและแก้ไขข้อบกพร่องเสียขั้นหนึ่งก่อนในรูปของ pretesting ซึ่งจะทำให้สามารถปรับปรุงแบบสอบถามให้ดีขึ้นได้โดยอาศัยผลการทดสอบ และถ้าในคำถามนั้นจะต้องใช้ศัพท์วิชาการ (technical terms) อะไรซึ่งอาจจะยากเกินไปสำหรับผู้ตอบก็ควรจะมีคำจำกัดความหรือคำอธิบายประกอบให้พอเพียงด้วย เช่น ในการสำรวจสถิติการเกษตรควรจะอธิบายให้ชัดเจนว่าเนื้อที่ถือครองหมายถึงอะไร เป็นต้น

(9) การคัดเลือกและฝึกอบรมพนักงานสำรวจและผู้ควบคุมการปฏิบัติงาน

(selection & training of interviewers & supervisors)

การคัดเลือกคนที่จะใช้ไปทำงานสนามในทั้งสองระดับควรจะต้องทำการทดสอบความเหมาะสมต่าง ๆ เช่น

- ก. ทดสอบเชาวน์และความเข้าใจเกี่ยวกับหนังสือ
- ข. ความถนัดเกี่ยวกับตัวเลข
- ค. ทดลองปฏิบัติงานสนามดูบ้างว่าเหมาะสมแก่งานแค่ไหน หรือไม่?

ถ้าเราไม่สามารถคัดเลือกพนักงานที่มีคุณสมบัติที่ดีหรือพอเชื่อถือได้พอสมควรแล้วคุณภาพของข้อมูลก็อาจจะไม่ดีเท่าที่ควรไม่ว่าการวางแผนในด้านอื่นจะดีอย่างไร ในกรณีนี้อาจมีปัญหาเรื่องการลงทะเบียนที่หรือคนเข้า ๆ ออก ๆ มาก ๆ ควรจะมีพนักงานสำรอง (reserved group) ไว้จำนวนหนึ่งเพื่อเข้าเสริมในกรณีทำงานอาจจะไม่สำเร็จตามกำหนด สำหรับด้านตัวผู้ควบคุมหรือ field supervisor นั้น ก็ควรวางแผนไว้ด้วย

เหมาะสมว่าจะใช้ผู้ควบคุมต่อนักงานกี่คนเพื่อให้สามารถทำหน้าที่กำกับและตรวจสอบดูแล การปฏิบัติงานของพนักงานสำรวจแต่ละคนได้อย่างมีประสิทธิภาพ พนักงานควบคุมหรือ supervisor นี้ควรจะได้รับ การฝึกอบรมในด้านหลักวิชาให้มากกว่าพนักงานสำรวจเพื่อ จะได้ช่วยวินิจฉัยและแก้ปัญหาเฉพาะหน้าต่าง ๆ ได้อย่างถูกต้องเหมาะสม

(10) การควบคุมคุณภาพของการปฏิบัติงานสนาม (quality control of field work) จากประสบการณ์ของผู้ที่เกี่ยวข้องกับการสำรวจข้อมูลสถิติ โดยการส่งเจ้าหน้าที่ออกไปสัมภาษณ์และบันทึกข้อมูลจากผู้ที่อยู่ในข่ายการสำรวจ อาจจะยืนยัน ได้ว่าคุณภาพของข้อมูลที่ได้จากการสำรวจ นอกจากจะขึ้นอยู่กับตัวประกอบสำคัญ ๆ คือ ระเบียบวิธีการสำรวจ (survey methodology) ความร่วมมือของผู้ที่อยู่ในข่ายจะต้อง ให้ข้อมูล (cooperation from respondents) คุณภาพของพนักงานแจงนับ (quality of enumerators) แล้วยังขึ้นอยู่กับประสิทธิภาพของการควบคุมงานสนาม (control of field work) ว่าทำได้แค่ไหน เพียงไร การควบคุมงานสนามในที่นี้หมายถึง การควบคุมดูแลให้พนักงานสำรวจปฏิบัติงานให้เป็นไปตามแผนงาน ทั้งในด้านปริมาณและ คุณภาพของการแจงนับ โดยให้มีกลไกสำหรับช่วยแก้ไขปัญหาเฉพาะหน้าต่าง ๆ เช่น การ แก้ปัญหาเรื่อง ไม่สามารถหาผู้ตอบหรือไม่ได้รับความร่วมมือจากผู้ตอบ (non-response problem) ซึ่งมักเป็นปัญหาใหญ่ที่พบเสมอในการสำรวจข้อมูลจากประชากรซึ่งเป็น human population นอกจากนี้ยังมีปัญหาในด้านการปฏิบัติงานล่าช้าหรือทำได้น้อยกว่าเป้าหมายที่ตั้งไว้ ซึ่งผู้กำกับงานสนามควรจะต้องหามาตรการหรือกลวิธีที่เหมาะสมในการแก้ไข ปัญหาดังกล่าว

(11) การตรวจสอบแบบกันทิกข้อมูลที่ได้รับคืนมา (inspection of returns) นับเป็นขั้นที่สำคัญอีกขั้นหนึ่งที่จะต้องทำนอกเหนือไปจากการควบคุมคุณภาพใน สนาม กล่าวคือในขณะที่พนักงานสำรวจยังปฏิบัติงานอยู่ในสนามควรจะได้มีการตรวจสอบ คุณภาพเบื้องต้น (initial quality check) เพื่อให้สามารถเพิ่มเติมรายการข้อมูลที่ยัง ไม่สมบูรณ์หรือขาดหายไป รวมทั้งการแก้ไขความไม่แน่นอนของข้อมูลที่เห็นได้ชัด (apparent inconsistencies) และเพื่อการนี้ในขั้นต่อมาควรจะได้มีเจ้าหน้าที่ชุดหนึ่งทำหน้าที่ตรวจทานรายการข้อมูลในใบแบบอย่างถี่ถ้วนก่อนการประมวลผลอีกขั้นหนึ่งด้วย

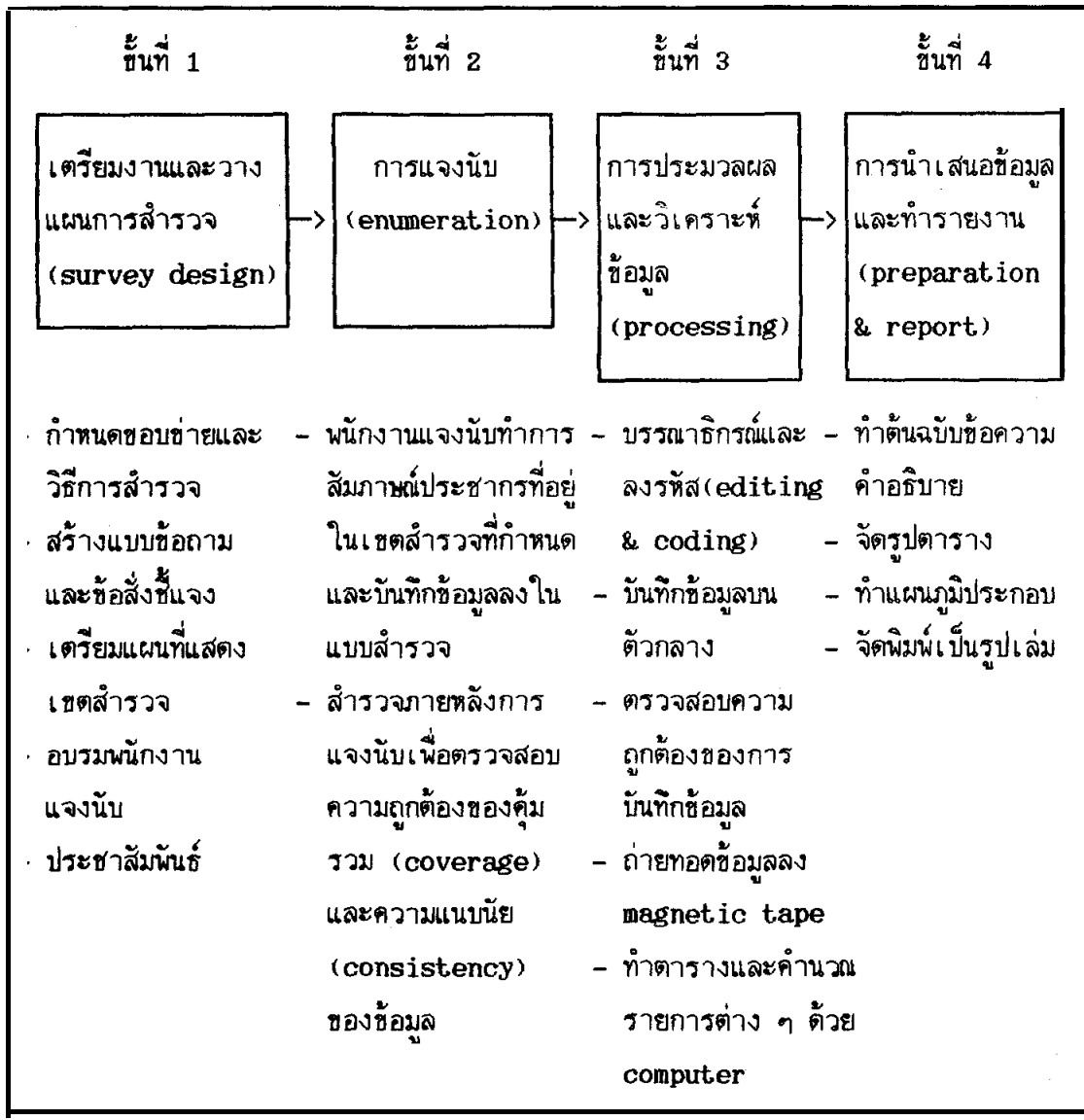
(12) การประมวลผลและวิเคราะห์ข้อมูล (processing & analysis of data) นับเป็นขั้นที่สำคัญอีกขั้นหนึ่งของการสำรวจเชิงสถิติ โดยทั่ว ๆ ไป

การประมวลผลข้อมูล (data processing) มักเกี่ยวข้องกับกระบวนการที่สำคัญๆ 3-4 อย่าง คือ การจำแนกและจัดหมวดหมู่ของข้อมูล (classifying & sorting) เช่น การจำแนกตามภาค จำแนกตามขนาดของเนื้อที่หรือลักษณะการถือครอง ฯลฯ การคำนวณ (calculating) เช่น คำนวณผลผลิต เปรียบเทียบต้นทุนการผลิต ฯลฯ การย่อหรือสรุปข้อมูล (summarizing) เช่น ย่อไว้ในรูปของตารางต่าง ๆ ซึ่งมีความหมายและสะดวกต่อการใช้

การประมวลผลดังกล่าวไม่ว่าจะทำด้วยมือ (manual processing method) หรือทำด้วยเครื่องจักร (machine processing method) จำเป็นจะต้องอาศัยการวางระบบงาน (system design) และควบคุมคุณภาพ (quality control) ที่ดี มิฉะนั้นแล้วความผิดพลาดคลาดเคลื่อนในขั้นประมวลผล (processing error) อาจจะมีมากจนกระทั่งให้ผลลัพธ์ที่คลาดเคลื่อนไปจากความเป็นจริงมาก ๆ ในยุคของการประมวลผลสมัยใหม่เราอาจจะมีเครื่องคอมพิวเตอร์ที่สามารถทำงานประมวลผลขั้นต่างๆ ได้อย่างอัตโนมัติและรวดเร็ว แต่ก็ควรจะต้องเข้าใจว่าคนจะต้องเป็นผู้วางแผนการใช้และจัดระบบการประมวลผลและเขียนคำสั่งแสดงรายละเอียดของแต่ละขั้นขึ้นก่อนคอมพิวเตอร์จึงจะทำการประมวลผลที่เราต้องการได้อย่างมีประสิทธิภาพ

(13) การจัดทำรายงานการสำรวจ (preparation of survey report) เมื่อทำการประมวลผลและวิเคราะห์ข้อมูลต่างๆจากการสำรวจแล้ว ขั้นสุดท้ายก็คือ การทำรายงานเพื่อเสนอผลการสำรวจเพื่อให้ผู้ที่สนใจจะใช้ข้อมูลได้ทราบผลลัพธ์ที่สำคัญๆ และน่าสนใจ การทำรายงานการสำรวจนี้ควรจะได้กล่าวถึงระเบียบวิธีการสำรวจ (survey methodology) ด้วย เพื่อให้ผู้ใช้ได้ทราบและพิจารณาว่าข้อมูลที่ได้มาควรมีความเชื่อถือได้ (reliability) แค่ไหน เพียงไร สำหรับผลลัพธ์ที่สำคัญๆ ของการสำรวจนอกจากจะเสนอไว้ในรูปตารางเชิงวิเคราะห์ (analytical tables) ต่าง ๆ แล้วควรมีแผนภูมิ (charts) ที่เหมาะสม แสดงผลดังกล่าวด้วยเพื่อประโยชน์ของผู้ที่ไม่สนใจอ่านตัวเลขและทำให้รายงานน่าสนใจมากขึ้นอีก เพราะคนจำนวนไม่น้อยไม่ชอบอ่านตารางตัวเลขแต่อาจจะสนใจดูแผนภูมิหรือกราฟมากกว่า

Diagram ต่อไปนี้แสดงขั้นที่สำคัญของงานสำรวจสถิติ (analytical survey operation) ที่จะต้องทำโดยทั่ว ๆ ไป ซึ่งอาจสรุปได้เป็น 4 ขั้น ดังนี้



การกำหนดแผนการสุ่มตัวอย่าง

การสำรวจด้วยตัวอย่าง (sample survey) จำเป็นที่จะต้องเลือกตัวอย่าง หน่วยงำนนั้นที่จะต้อง เป็นตัวแทนของหน่วยงำนนั้นในคุ่มรวม การงำนนั้นจะทำเฉพาะใน หน่วยงำนนั้นตัวอย่างเท่านั้น แต่ข้อมูลสถิติที่แสดงจะเป็นยอดข้อมูลสถิติของคุ่มรวมทั้งหมด ยอดข้อมูลนั้นได้มาจากการประมาณโดยสูตรทางคณิตศาสตร์

การกำหนดแบบแผนของตัวอย่าง (sample design) ที่จะใช้ในการสำรวจ เป็นส่วนสำคัญที่สุดส่วนหนึ่งของการรวบรวมสถิติ แบบแผนตัวอย่างควรจะเป็นแผนที่ให้ ได้ตัวอย่างมา โดยหลักเกณฑ์ทาง ทฤษฎีความน่าจะเป็น ไม่ใช่เป็นการเจาะจงเลือกตัวแทนที่ผู้เลือก เข้าใจว่าเป็นตัวแทนที่ดีโดยวิธีอัตนัย เพราะหากเป็นเช่นนั้นผู้เลือกอาจมีความลำเอียงเลือกตัวแทนที่ไม่ใช่ตัวแทนที่แท้จริงของหน่วยงำนนั้น หรืออาจจะเลือกโดยความเชื่อของตัวเองอย่างสุจริตใจว่าหน่วยงำนนั้นมีลักษณะใดลักษณะหนึ่งจะเป็นตัวแทนที่ดี ซึ่งในความเป็นจริงหาเป็นเช่นนั้นไม่

แก่นสารสำคัญของตัวอย่างที่เลือกโดยทฤษฎีความน่าจะเป็นคือ หน่วยงำนนั้น หน่วยใดหน่วยหนึ่งจะมีความน่าจะเป็นที่จะถูกสุ่มออกมา ซึ่งความน่าจะเป็นนี้ต้องไม่เท่ากับ 0 และนักสถิติผู้เลือกหน่วยตัวอย่างจะต้องรู้ความน่าจะเป็นในการสุ่มก่อนที่จะทำการสุ่มตัวอย่างที่เลือกมาได้ โดยให้ความน่าจะเป็นในการเลือกเท่ากันหมดในทุกหน่วยงำนนั้น มีชื่อเรียกเป็นที่รู้จักกันทั่วไปว่าเป็นการสุ่มแบบธรรมดาหรือแบบง่าย (simple random sample)

ตัวอย่างที่สุ่มโดยทฤษฎีความน่าจะเป็นเท่านั้นที่ถือว่าเป็นตัวอย่างที่เป็นตัวแทนที่ดีของหน่วยงำนนั้นในคุ่มรวม การประมาณยอดสถิติก็จะทำได้โดยสูตรทางคณิตศาสตร์ ความคลาดเคลื่อนจากการประมาณ ก็สามารถคำนวณได้จากสูตรทางคณิตศาสตร์ ทำนองเดียวกับปริมาณความเอียงเจจากการประมาณ (หากจะมี) ก็สามารถคำนวณได้

ตัวอย่างที่เลือกโดยทฤษฎีความน่าจะเป็นเองก็มีแบบแผนของการสุ่มหลายแบบด้วยกัน และมีชื่อเรียกต่าง ๆ กัน ในทางปฏิบัติแล้วการสำรวจหนึ่ง ๆ ก็ใช้แบบแผนการสุ่มไปอย่างหนึ่ง โดยไม่ซ้ำกัน ในชั้นวางแผนการสำรวจด้วยตัวอย่าง นักสถิติจะต้องเตรียมแบบแผนการสุ่มตัวอย่าง และต้องทำการสุ่มเพื่อให้ได้ตัวอย่างหน่วยงำนนั้นให้พร้อมก่อน การออกงำนนั้น

ข้อจะต้องพิจารณาในขณะที่เตรียมตัวอย่าง

โดยทั่ว ๆ ไปนักสถิติจะต้องพยายามใช้แผนแบบการสุ่มตัวอย่างที่จะทำให้ข้อมูลที่ต้องการจะรวบรวมมีคุณภาพดีที่สุด กล่าวคือ (1) ต้องเป็นข้อมูลที่ไม่มีความเอียงเอน หรือหากมีความเอียงเอน ก็มีแต่เพียงเล็กน้อย (2) ข้อมูลต้องมีความคลาดเคลื่อนจากการประมาณน้อยที่สุด (3) การปฏิบัติงานเก็บรวบรวมตามแผนแบบตัวอย่างที่กำหนดอยู่ในวิธีที่จะทำได้โดยไม่ยากนัก (4) ค่าใช้จ่ายในการดำเนินงานทั้งหมดก็ต้องอยู่ในวงเงินที่กำหนดไว้

ประการแรก ข้อมูลจะไม่มีความเอียงเอนได้ก็ต่อเมื่อ ตัวอย่างที่ใช้นั้นเลือกมาได้โดยทฤษฎีความน่าจะเป็น และโดยการใช้สูตรของการประมาณที่เหมาะสม

ประการที่สอง การที่จะให้ข้อมูลที่ประมาณได้มีความคลาดเคลื่อนน้อยที่สุดภายในวงเงินที่กำหนดให้และให้ได้ตัวอย่างที่สะดวกกับการทำงาน นักสถิติจะต้องพิจารณาข้อประเด็นดังนี้ :-

(ก) กำหนดหน่วยสุ่มตัวอย่าง (sampling unit) ที่เหมาะสม หน่วยแฉงนับคือหน่วยที่พนักงานไปสัมภาษณ์และบันทึกข้อมูลเบื้องต้น หน่วยสุ่มตัวอย่างอาจจะเป็นหน่วยเดียวกับหน่วยแฉงนับก็ได้ หรืออาจจะไม่ใช่ก็ได้ เช่น ในการสำรวจจำนวนคนเกิด/ตาย เราอาจจะใช้บ้านเป็นหน่วยแฉงนับ ซึ่งมีประมาณ 30,000 บ้านในตัวอย่าง แต่ในการสุ่มตัวอย่างเราอาจจะไม่สุ่มบ้าน 30,000 บ้านโดยตรง แต่สุ่มหมู่บ้านออกมา 300 หมู่บ้าน (1 หมู่บ้านมีประมาณ 100 บ้านโดยประมาณ) โดยถือว่าเมื่อสุ่มได้หมู่บ้านใดเป็นหมู่บ้านตัวอย่างแล้ว บ้านทุกบ้านในหมู่บ้านถือว่าเป็นบ้านที่ตกอยู่ในตัวอย่างและอยู่ในขอบข่ายที่จะถูกแฉงนับ ในกรณีนี้หน่วยสุ่มตัวอย่างคือหมู่บ้าน แต่หน่วยแฉงนับคือบ้าน

จากตัวอย่างที่ให้ ให้เราวาดภาพการแฉงนับ 30,000 บ้าน กระจายไปแทบทุกหมู่บ้านทั่วราชอาณาจักรเปรียบเทียบกับการแฉงนับ 30,000 บ้าน ภายใน 300 หมู่บ้าน การทำงานด้วยหน่วยแฉงนับแบบหลังจะสะดวกกว่าแบบแรกมาก พนักงานไม่ต้องเสียค่าเดินทางและเสียเวลาเดินทางมากประการหนึ่ง และไม่ต้องเสียเวลาตระเวนหาบ้านตัวอย่าง เพราะบ้านตัวอย่างอยู่รวมกันเป็นกลุ่มในหมู่บ้านตัวอย่างแล้ว การควบคุมพนักงานก็ทำได้ง่ายกว่า เพราะรู้แน่ ๆ ว่าพนักงานทำงานในพื้นที่ภายในหมู่บ้านไหนเป็นการแน่นอน ค่าใช้จ่ายในกรณีหลัง จะตกบ้านละประมาณ 4 บาท แต่ในกรณีแรกค่าใช้จ่ายในการแฉงนับจะตกประมาณบ้านละ 25 บาท (ค่าใช้จ่ายในปี 2510)

การกำหนดหน่วยสุ่มตัวอย่าง จึงมีส่วนสำคัญที่สุดต่อค่าใช้จ่ายในการสำรวจ และต่อแผนการปฏิบัติงานในชั้นเจเนนบ

(ข) ข้อมูลที่ใช้เป็นในการสุ่มตัวอย่าง ก่อนสุ่มตัวอย่างใด ๆ อย่างน้อยที่สุดจะต้องมี เอกสารแสดงรายชื่อของหน่วยสุ่มตัวอย่างอย่างครบถ้วนภายในคุ่มรวม จึงจะทำการสุ่มตัวอย่างได้ รายชื่อดังกล่าวเราเรียกว่า กรอบตัวอย่าง (sampling frame) ในชั้นเตรียมงาน นักสถิติจะต้องพิจารณาว่ารายชื่อของหน่วยสุ่มตัวอย่างที่จะใช้จะหามาได้โดยง่ายหรือไม่ และจะต้องสิ้นงบประมาณเท่าใด จะถูกต้องตามคุ่มรวมที่เป็นจริงหรือไม่

ในการสำรวจจำนวนคนเกิด/ตายที่เราพูดถึง หากจะใช้บ้านเป็นหน่วยสุ่มตัวอย่าง เราจะต้องเตรียมกรอบตัวอย่างประกอบด้วยบัญชีรายชื่อเจ้าบ้านและเลขที่บ้านประมาณ 6,000,000 บ้านก่อนที่จะทำการสุ่ม กรอบตัวอย่างนี้จะเตรียมมาได้ก็ต่อเสียเงินค่าไปทำการบันทึกนับจดบ้านทั่วราชอาณาจักร อาจจะมีค่าใช้จ่ายเพื่อการนี้ประมาณ 3,000,000 บาท แต่หากเราจะใช้หมู่บ้านเป็นหน่วยสุ่มตัวอย่าง กรอบตัวอย่างคือรายชื่อหมู่บ้าน ประมาณ 50,000 หมู่บ้าน ในประเทศไทยมีอยู่แล้วอย่างครบถ้วน จะขอมาใช้ได้จากกระทรวงมหาดไทยโดยไม่ต้องเสียค่าใช้จ่ายเลย เป็นต้น

(ค) ข้อมูลอื่น ๆ ที่เกี่ยวข้องที่อาจจะนำมาใช้ประกอบการวางแผนแบบการสุ่มตัวอย่าง ให้มีประสิทธิภาพและใช้ประกอบในการประมาณข้อมูลดังกล่าวอาจจะนำมาใช้ได้หลายกรณี เช่น ดังต่อไปนี้ :-

(1) คำนวณล่วงหน้าว่าความคลาดเคลื่อนจากการสุ่มตัวอย่างจะเป็นเท่าไร สำหรับขนาดตัวอย่างที่กำหนดให้ ในทางกลับกันสำหรับประมาณความคลาดเคลื่อนที่กำหนดให้การคำนวณจะชี้ให้เห็นว่าขนาดตัวอย่างควรจะเป็นเท่าไร? ความคลาดเคลื่อนจะทำให้ลดลงให้เหลือเท่าใดก็ได้ ด้วยการขยายขนาดของตัวอย่างจนพอเพียง ประมาณอย่างคร่าว ๆ เมื่อขยายตัวอย่างเป็น 4 เท่า ความคลาดเคลื่อนจะลดลงครึ่งหนึ่ง

(2) นำข้อมูลที่เกี่ยวข้องมาแบ่งหน่วยเจเนนบออกเป็นชั้นภูมิก่อนทำการสุ่ม และใช้แผนแบบการสุ่มจากแต่ละชั้นภูมิ (Stratified Sampling) โดยกำหนดขนาดตัวอย่างภายในแต่ละชั้นภูมิที่เหมาะสม ตัวอย่างที่สุ่มมาได้ด้วยวิธีนี้อาจจะมีประสิทธิภาพสูงกว่าตัวอย่างสุ่มโดยไม่แบ่งชั้นภูมิ

(3) นำข้อมูลที่เกี่ยวข้องมาคำนวณความน่าจะเป็นที่จะสุ่มหน่วยตัว-

อย่างใด ๆ เช่น ในการสุ่มหมู่บ้านเพื่อการสำรวจการเกิด/ตาย เราอาจจะใช้ข้อมูลจำนวนบ้าน (จากทะเบียนราษฎร์) เป็นเกณฑ์กำหนดให้ความน่าจะเป็นที่จะสุ่มแต่ละหมู่บ้านเป็นปฏิภาคโดยตรง (directly proportional-to) กับจำนวนบ้านจากทะเบียนราษฎร์ ตัวอย่างที่สุ่มโดยวิธีนี้อาจจะมีประสิทธิภาพสูงกว่าตัวอย่างสุ่มโดยวิธีธรรมดา

(4) นำข้อมูลที่เกี่ยวข้องมาใช้ประกอบในสูตรการประมาณ เช่น ในการประมาณแบบอัตราส่วน (ratio estimate) หากเลือกข้อมูลที่มาช่วยการประมาณได้อย่างเหมาะสม ความคลาดเคลื่อนของข้อมูลสถิติที่ประมาณได้จะลดลงได้อย่างมาก

ประการที่สาม นักสถิติจะต้องพิจารณาว่ามีความจำเป็นจะต้องแจงนับทุกหน่วยแจงนับในหน่วยสุ่มตัวอย่างหรือไม่ เช่น ในการสำรวจคนเกิด/ตาย หากหมู่บ้านเป็นหน่วยสุ่มตัวอย่าง นักสถิติจะต้องพิจารณาว่าจะต้องแจงนับทุกบ้านในหมู่บ้านหรือไม่ หรือจะแจงนับเพียงบางบ้านในหมู่บ้าน ซึ่งบางบ้านนั้นก็จะได้มาจากการสุ่มอีกครั้งหนึ่งภายในหมู่บ้าน ตัวอย่างนั่นเอง จะเห็นว่าในกรณีนี้จะมีการสุ่ม 2 ครั้ง คือ สุ่มหมู่บ้านครั้งหนึ่ง และสุ่มบ้านภายในหมู่บ้านตัวอย่างอีกครั้งหนึ่ง มีชื่อเรียกตัวอย่างประเภทนี้ว่า ตัวอย่างแบบ 2 ชั้น (two-stage sample) หากมีการสุ่มในชั้นที่ 2 เกิดขึ้น นักสถิติก็ต้องวินิจฉัยว่าควรจะใช้ขนาดตัวอย่างที่บ้านภายในแต่ละหมู่บ้านตัวอย่าง

การสุ่มตัวอย่างไม่จำเป็นจะต้องจำกัดลงไปว่าสุ่มเพียง 2 ชั้น แต่อาจจะขยายออกไปยังชั้นที่ 3 ชั้นที่ 4 ต่อ ๆ ไปได้ เช่น ในการสำรวจผลผลิตผลการเกษตร ซึ่งสมมติว่าขนาดตัวอย่าง = 20,000 ครัวเรือนเกษตร เราอาจจะใช้หน่วยสุ่มตัวอย่างในแต่ละชั้นดังนี้

สุ่มชั้นที่ 1 ใช้หมู่บ้านเป็นหน่วยสุ่มตัวอย่าง สุ่ม 2,000 หมู่บ้าน

สุ่มชั้นที่ 2 สุ่มผู้ประกอบการเกษตร หมู่บ้านละ 10 ผู้ประกอบการ ได้ตัวอย่าง 20,000 ผู้ประกอบการ เพื่อเก็บข้อมูลเบื้องต้นจากการสัมภาษณ์ และวัดเนื้อที่ถือครองทั้งหมด

สุ่มชั้นที่ 3 สุ่มแปลงเก็บเกี่ยวขนาด 5m x 5m จากเนื้อที่ทั้งหมดของผู้ประกอบการเกษตรตัวอย่าง โดยสุ่มผู้ประกอบการละ 3 แปลง รวมเป็นขนาดตัวอย่างแปลงเก็บเกี่ยวทั้งหมดรวม 1,500,000 ตารางเมตร

ตัวอย่างที่ได้จะเป็นตัวอย่างแบบ 3 ชั้น (three-stage sample) ตัวอย่างในชั้นที่สอง เราเรียกว่าตัวอย่างย่อย (sub-sample) ของชั้นที่ 1 และตัวอย่างใน

ขั้นที่ 3 เราเรียกว่า ตัวอย่างย่อยของขั้นที่ 2 ในการสุ่มแต่ละชั้นอาจจะใช้ความน่าจะเป็นในการสุ่มต่าง ๆ กันก็ได้ นักสถิติผู้วางแผนแบบการสุ่มตัวอย่างจะต้องวินิจฉัยต่อไปด้วยว่า หากมีการสุ่มตัวอย่างแบบหลายชั้นนี้ควรจะใช้ขนาดตัวอย่างในแต่ละชั้นอย่างไร เช่น หากจะใช้ตัวอย่าง 20,000 ผู้ประกอบการ เราควรจะต้องเลือก 2,000 หมู่บ้าน และใช้ตัวอย่างย่อยหมู่บ้านละ 10 บ้าน หรือควรจะต้องเลือกเพียง 1,000 หมู่บ้านและใช้ตัวอย่างย่อยหมู่บ้านละ 20 บ้าน เป็นต้น

ข้อกำหนดดังกล่าวเป็นส่วนหนึ่งของแผนแบบการสุ่มตัวอย่าง นักสถิติจะต้องวินิจฉัยโดยอาศัยข้อมูลที่เกี่ยวข้องที่มีอยู่ รู้จักเอาข้อมูลเหล่านั้นมาวิเคราะห์เพื่อตัดสินใจเลือกแผนแบบการสุ่มตัวอย่างที่มีประสิทธิภาพที่สุด

ในขั้นเตรียมงาน นักสถิติไม่แต่เพียงจะต้องกำหนดแผนแบบตัวอย่างเท่านั้น แต่จะต้องเตรียมกรอบตัวอย่าง ซึ่งในการนี้อาจจะต้องทำการคัดลอกข้อมูลที่จะใช้จากแหล่งต่าง ๆ บางครั้งเอกสารเหล่านี้ได้มาจากการนับจคนในสนามเพิ่มเติม หรือจากการรายงานของผู้ใหญ่บ้าน ต้องทำการสุ่มจนได้ตัวอย่างครบตามต้องการ และต้องเตรียมเอกสารแสดงตำบลที่ตั้งของหน่วยตัวอย่าง เพื่อพนักงานแจกนับจะ ได้ออกไปรวบรวมจากหน่วยแจกนับภายในตัวอย่างได้

ท้ายที่สุดนักสถิติผู้วางแผนแบบการสุ่มตัวอย่าง จะต้องเตรียมสูตรในการประมาณข้อมูลสถิติไว้ให้พร้อม พร้อมทั้งสูตรที่จะใช้ในการคำนวณความคลาดเคลื่อนจากการประมาณด้วย สูตรดังกล่าวจะต้องสอดคล้องกับแผนแบบการสุ่มตัวอย่างที่กำหนดขึ้น

การประชาสัมพันธ์

โครงการเก็บข้อมูลสถิติใหญ่ ๆ เช่น สำมะโนหรือสำรวจในระดับชาติ ข้อมูลเบื้องต้นได้มาจากการสัมภาษณ์หัวหน้าครัวเรือน เจ้าของผู้ประกอบการ ฯลฯ คุณภาพของข้อมูลขึ้นอยู่กับความสามารถ ความพร้อมและเต็มใจของผู้ตอบที่จะให้คำตอบแก่พนักงานแจกนับ

งานเก็บข้อมูลสถิติ โครงการใหญ่ในระดับชาติจึงมีความจำเป็นที่จะต้องทำการประชาสัมพันธ์ให้ผู้ตอบทราบล่วงหน้าว่ารัฐกำลังจัดทำโครงการเก็บรวบรวมข้อมูลสถิติ และขอให้ผู้ตอบเตรียมคำตอบไว้ล่วงหน้า และให้ความร่วมมือแก่พนักงานเมื่อเวลาไปแจกนับ

ข้อมูลพื้นฐานที่บางรายการ อาจจะพาดพิงไปถึงเรื่องส่วนตัวของผู้ตอบ เช่น เกี่ยวกับรายได้หรือเรื่องธุรกิจการค้าหรืออาจจะเป็นเรื่องเกี่ยวกับสถานภาพการสมรสของ

ผู้ตอบเป็นต้น ตามปกติประชาชนมักจะหวาดระแวง เมื่อมีบุคคลแปลกหน้าเข้ามาซักถามและบันทึกข้อเท็จจริงต่าง ๆ เพราะกลัวว่าอาจจะกลายเป็นสายลับของผู้ร้าย ในบ้านของตน ยิ่งกว่านั้นผู้ตอบอาจจะมีความเกรงกลัวว่า เมื่อตอบไปแล้วอาจจะมีผลสะท้อนนำความผิดพลาดมาสู่ตนได้ (เช่น เรื่องเกี่ยวกับรายได้ซึ่งตนได้ บิดบังสรรพากรไว้เพื่อเลี่ยงภาษี หากบอกความจริงแล้วอาจจะรื้อไต่ไปถึงสรรพากรได้) หรือหากเป็นธุรกิจการค้าก็เกรงว่าเมื่อตอบความจริง ความลับอาจจะรื้อไต่ไปถึงคู่แข่งทำให้เกิดการเสียเปรียบในเชิงการค้าได้ ประเด็นสำคัญที่การประชาสัมพันธ์จะต้องเน้นก็คือ การเก็บรวบรวมข้อมูลสถิตินี้เพื่อ **การสถิติเท่านั้น** กล่าวคือ ข้อมูลเบื้องต้นรายบุคคลจะไม่นำไปเปิดเผยเพื่อการใดเป็นอันขาด ซึ่งความจริงมีพระราชบัญญัติคุ้มครองผู้ตอบอยู่ หากมีผู้นำไปเปิดเผย ผู้นั้นต้องได้รับโทษถึงจำคุก ในยอดข้อมูลสถิติที่ประมวลมาแล้วและเผยแพร่จะไม่มีใครทราบข้อมูลของแต่ละบุคคลมีรายละเอียดเป็นอย่างไร

ผู้ตอบมักจะไม่ใช่ใจว่าข้อมูลเบื้องต้นที่ตัวเองตอบไปแล้วจะนำผลประโยชน์อันใดมาสู่ตนได้บ้าง ในเรื่องนี้การประชาสัมพันธ์จะต้องหาวิธีอธิบายให้ประชาชนได้ทราบในภาษาที่จะเข้าใจได้ง่าย ๆ โดยไม่สาธยายถึงประโยชน์จนเกินความเป็นจริง

ส่วนวิธีการประชาสัมพันธ์นั้น ก็คงใช้สื่อมวลชนประเภทต่าง ๆ อาจจะโดยวิธีให้หัวหน้าโครงการสัมภาษณ์ทางหนังสือพิมพ์บ้าง พิมพ์ใบปลิวแจกบ้าง ทำแผ่นโปสเตอร์ติดบ้าง ทำแผ่นสไลด์ไปฉายตามโรงหนังบ้าง ประกาศโดยรถโฆษณาบ้าง ทางวิทยุ หรือทางโทรทัศน์บ้าง หรืออาจจะจัดการบรรยายชี้แจงให้แก่ประชาชนในที่ชุมชน หรือในท้องประชุม หรือในบางครั้งอาจจะทำแสดมภ์ หรือเหรียญตราเป็นที่ระลึก โครงการเก็บข้อมูลก็ได้

สำหรับในประเทศไทยนั้น กรณีที่การเก็บข้อมูลต้องทำในหมู่บ้าน การประชาสัมพันธ์ที่ที่สุดและเสียค่าใช้จ่ายน้อยที่สุด ก็คือทำผ่านสายการปกครองมหาดไทย เริ่มด้วยการประชาสัมพันธ์ให้ผู้ว่าราชการจังหวัด นายอำเภอ ได้ทราบ นายอำเภอชี้แจงในชั้นต่อไปให้กำนัน ผู้ใหญ่บ้าน ได้ทราบ และให้ผู้ใหญ่บ้าน เรียกลูกบ้านประชุมเพื่อชี้แจงอีกทีหนึ่ง อาจจะต้องเตรียมค่าประกาศจ่าย ๆ ให้เป็นคู่มือให้ผู้ใหญ่บ้านได้ชี้แจงโดยไม่ผิดไปจากความเป็นจริง

ชนิดของตัวอย่าง

ตัวอย่างมี 3 ชนิด คือ

1. ตัวอย่างที่มาจากการใช้ทฤษฎีความน่าจะเป็น เรียกว่า **probability samples**
2. ตัวอย่างจากดุลยพินิจของผู้สุ่มตัวอย่าง เรียกว่า **judgement samples**
3. ตัวอย่างที่ได้มาโดยถือความสะดวก เรียกว่า **convenience samples**

Probability Samples

คือตัวอย่างซึ่งเกิดจากการคัดเลือก จากประชากรโดยวิธีความน่าจะเป็น การเลือกตัวอย่างจากประชากรโดยใช้ความน่าจะเป็น คือการเลือกตัวอย่าง โดยให้สมาชิกทุกหน่วยมีสิทธิ์ หรือโอกาสเท่าเทียมกันที่จะปรากฏเป็นตัวอย่าง วิธีนี้มีข้อดี 2 ข้อ คือ

1. ข้อมูลที่ได้จากตัวอย่าง สามารถใช้วิธีการทางสถิติกำหนดความคลาดเคลื่อนจากการสุ่มตัวอย่าง ได้
2. เป็นการหลีกเลี่ยงความอคติ (biases) อันเกิดจากการใช้ดุลยพินิจการตัดสินใจ ของผู้สุ่มตัวอย่าง

ตัวอย่างที่ใช้ความน่าจะเป็นมีประโยชน์ และสำคัญที่สุดในการใช้อ้างอิงเชิงสถิติ

Judgement Sample

เป็นตัวอย่างที่มีลักษณะตรงข้ามกับตัวอย่างความน่าจะเป็น ตัวอย่างแบบนี้จะ ได้มาจากการตัดสินใจ หรือดุลยพินิจในการถูกเลือก "เป็นตัวแทน" ของสมาชิกของประชากรนั้น

ตัวอย่างเช่นการกำหนด "ความเป็นตัวแทน" ของประชากร ของตลาดวงซึก-ฟอกชนิดหนึ่ง ผู้จัดการอาจกำหนด จังหวัดที่มีลักษณะ "พื้นฐาน" (typical) เพียงไม่กี่จังหวัดเพื่อทดสอบวงซึกฟอกสูตรใหม่ เช่น อาจเลือก กรุงเทพฯ เชียงใหม่ อุตรดิตถ์ นครศรีธรรมราช ชลบุรี พิษณุโลก สำหรับเป็นตลาดทดลอง ข้อมูลที่ได้จากตัวอย่างประเภทนี้จะใช้วิธีการทางสถิติหาความคลาดเคลื่อนจากการสุ่มตัวอย่างไม่ได้ บางครั้งการสุ่มตัวอย่าง โดยวิธีนี้ให้ผลลัพธ์แม่นยำมาก ถ้าหน่วยที่ตกเป็นตัวอย่าง ไม่มีลักษณะ "ความเป็นตัวแทน" ที่ดีของประชากรนั้น

การสุ่มตัวอย่างแบบโควตา (quota sample) เป็นลักษณะตัวอย่างชนิดหนึ่งของ judgement sample ด้วยวิธีกำหนดจำนวนตัวอย่างให้ผู้สัมภาษณ์ โดยกำหนดลักษณะ

ร่วมของตัวอย่าง เช่น เพศ อายุ ระดับความรู้ รายได้ และท้องที่อยู่อาศัย เป็นต้น ส่วนการเลือกตัวอย่างให้เป็นหน้าที่ของผู้สัมภาษณ์ ขอแต่เพียงให้ได้ลักษณะที่กำหนด ครบตามโควตาที่กำหนดให้ เท่านั้น ทั้งนี้โดยหวังว่า ผู้สัมภาษณ์จะใช้ดุลยพินิจที่ดีในการคัดเลือกตัวอย่างที่เหมาะสม แต่ในทางปฏิบัติส่วนใหญ่พบว่าผู้สัมภาษณ์มักเลือกตัวอย่างจากสิ่งใกล้ตัว คือญาติพี่น้อง หรือจากสิ่งที่พอจะหาได้ (available) ในเวลาที่ออกสัมภาษณ์ เช่น พบแต่แม่บ้านซึ่งไม่ต้องทำงานนอกบ้าน ตัวอย่างแบบนี้จึงมีความอคติหรือความเอียงเอน ซึ่งถ้าใช้วิธีความน่าจะเป็นจะไม่มีโอกาสเกิดขึ้น เช่น ถ้าประชากรที่อยู่ในความสนใจ คือผู้ซื้อรถยนต์ใหม่ของบริษัท บริษัทได้ส่งแบบสอบถามไปให้ลูกค้าที่ซื้อรถยนต์ใหม่ทั้งหมด แต่สมมติว่ามีผู้ตอบรับหรือส่งแบบสอบถามกลับคืนมาเพียง 50% ถ้าผู้วิจัยตกลง (สรุป) ว่า ผู้ตอบรับ 50% นี้ มีลักษณะพื้นฐานของผู้ซื้อใหม่ทั้งหมด เช่น ความเป็นตัวแทนด้านอายุ รายได้ ท้องที่อยู่อาศัย เป็นต้น ตัวอย่างนี้จึงเป็น "judgement sample" แม้ว่าเริ่มแรกจะใช้วิธีความน่าจะเป็นแต่เมื่อผู้วิจัยใช้เหตุผลตัดสินใจ เลือกผู้ซื้อ 50% นี้เป็น "ตัวแทน" ของผู้ซื้อทั้งหมด จึงกลายเป็นตัวอย่างแบบ judgement

Convenience Sample

ตัวอย่างแบบ convenience ไม่ใช่ตัวอย่างแบบความน่าจะเป็น และไม่เหมือนตัวอย่างแบบ judgement เพราะไม่มีความพยายามที่จะให้เกิด "ความเป็นตัวแทน" บางครั้งเรียกว่า "Chunk" คือแทนกลุ่มหนึ่ง จากประชากรซึ่งได้มาโดยความสะดวก เช่น ครูเลือกนักเรียนในชั้นเรียนของตนเพื่อศึกษาอิทธิพลของแรงจูงใจ ชั้นเรียนที่ครูสอนเป็นตัวอย่างแบบสะดวก (convenience sample) จะเห็นว่าครูไม่ได้คำนึงถึง "ความเป็นตัวแทน" ของประชากรที่อยู่ในความสนใจ อีกตัวอย่างคือ การเลือกตัวอย่างโดยอาศัยความสะดวก โดยเลือกจากญาติพี่น้องเพื่อนฝูงที่อยู่ใกล้ตัว จะเห็นว่าลักษณะตัวอย่างแบบ judgement และแบบ convenience คล้ายกันมาก ข้อแตกต่างคือ ตัวอย่างแบบ convenience ใช้ได้ สำหรับจุดประสงค์ที่จำกัด เพราะไม่มั่นใจว่าเป็นตัวแทนของประชากรที่อยู่ในความสนใจ

การสุ่มตัวอย่างแบบง่ายจากประชากรแบบจำกัด

(Simple Random Sampling From a Finite Population)

การสุ่มวิธีนี้เป็นวิธีหลักของตัวอย่างความน่าจะเป็น และมักเขียนสั้น ๆ ว่าตัวอย่างสุ่ม (random sample)

นิยาม

ตัวอย่างสุ่มแบบง่ายจากประชากรแบบจำกัด คือตัวอย่างที่เกิดจากวิธีการเลือกสรรโดยให้กลุ่มตัวอย่างที่เป็นไปได้ทุก ๆ กลุ่มในประชากรนั้นมีโอกาสโดยเท่ากัน ที่จะได้รับเลือกเป็นตัวอย่าง

เช่น มีนักศึกษาปริญญาเอก 5 คน คือ A, B, C, D, E ถ้าจะเลือกมาเป็นตัวอย่าง 2 คน จะมีกลุ่มตัวอย่างขนาด 2 คนที่เป็นไปได้ทั้งหมดในประชากรนั้นอยู่ 10 กลุ่มตัวอย่าง คือ

A, B	A, D	B, C	B, E	C, E
A, C	A, E	B, D	C, D	D, E

ถ้าจะใช้การสุ่มตัวอย่างแบบง่ายเพื่อเลือกนักเรียนตัวอย่างมา 2 คน จะต้องใช้วิธีการที่ทุกกลุ่มตัวอย่างมีโอกาสเท่ากัน คือ $.10$ ที่จะถูกเลือกเป็นตัวอย่าง

ข้อสังเกต

การสุ่มตัวอย่างนักเรียน 2 คน จากประชากร 5 คน ในตัวอย่างนั้น เป็นการสุ่มแบบไม่มีการแทนที่ ซึ่งเปิดโอกาสให้สมาชิกแต่ละหน่วยตกเป็นตัวอย่างได้เพียงครั้งเดียว ส่วนการสุ่มแบบมีการแทนที่ จะเปิดโอกาสให้สมาชิกหน่วยเดิมมีโอกาสตกเป็นตัวอย่างได้มากกว่า 1 ครั้ง ซึ่งการสุ่มแบบมีการแทนที่นี้ไม่ค่อยนิยมใช้กับประชากรแบบจำกัด การให้ได้มาซึ่งตัวอย่างแบบง่าย (การเลือกหรือหยิบตัวอย่าง)

การเลือกสมาชิกสัก หน่วย โดยการสุ่มแบบง่าย และไม่มีการแทนที่ จากประชากรแบบจำกัด N หน่วย จะมีขั้นตอน ดังนี้

1. เลือกหน่วยที่ 1 โดยให้ทุกหน่วยจาก N หน่วยมีโอกาสถูกเลือกเท่ากัน นั่นคือ มีโอกาส $1/N$ ที่จะถูกเลือกเป็นตัวอย่าง
2. เลือกหน่วยที่ 2 โดยให้ทุกหน่วยที่เหลือ คือ $N-1$ หน่วยมีโอกาสเท่ากันที่จะถูกเลือก นั่นคือมีโอกาส $1/(N-1)$ ที่จะถูกเลือกเป็นตัวอย่าง

3. ทำซ้ำเช่นนี้กับหน่วยต่อไปเรื่อย ๆ จนได้ครบ n ตัวอย่าง
สิ่งสำคัญอีกอันหนึ่งของการสุ่มตัวอย่างแบบง่ายคือ ต้องมี **กรอบตัวอย่าง**

นิยาม

กรอบตัวอย่าง คือรายชื่อสมาชิกทุกหน่วยของประชากร

เช่นต้องการสุ่มนักศึกษาตัวอย่าง กรอบตัวอย่างคือรายชื่อนักศึกษารหัสต่าง ๆ
ทุกรหัส ที่ฝ่ายทะเบียน (โดยเครื่องคอมพิวเตอร์) รวบรวมไว้

ถ้ากรอบตัวอย่างไม่สมบูรณ์ เช่นต้องการรายชื่อหน่วยความทั่วประเทศ ถ้าเรา
ใช้รายชื่อจากสมาคมหน่วยความอาจไม่ครอบคลุมหน่วยทั่วประเทศ แต่เนื่องจากเป็นแหล่ง
ข้อมูลที่ดีที่สุด จึงต้องใช้รายชื่อนี้ เช่นนี้จะเกิดประชากร 2 อย่างคือ ประชากรเป้าหมาย
(target population) ในที่นี้คือ หน่วยความทั่วประเทศ และประชากรกลุ่มตัวอย่าง
(sampled population) ในที่นี้คือ หน่วยความที่เป็นสมาชิกของสมาคมหน่วยความ ใน
กรณีที่กรอบตัวอย่างของประชากร 2 แบบนี้แตกต่างกัน ควรใช้ความพยายามอย่างมากที่สุด
ให้กรอบตัวอย่าง 2 อันนี้ ซ้ำซ้อนกันมากที่สุด จะได้ไม่มีข้อผิดพลาดมาก

การใช้ตารางเลขสุ่ม

ตาราง B-8 เป็นตัวอย่างของตารางเลขสุ่ม เพื่อใช้สำหรับเลือกตัวอย่างสุ่ม
แบบง่าย ตารางเลขสุ่มประกอบด้วยเลข 0, 1, 2, ..., 9 ซึ่งประกอบด้วยความน่า-
จะเป็นเท่ากัน คือ .10 จึงเปิดโอกาสให้เลือกใช้เลขจาก 00 ถึง 99 หรือ 000 -
999 หรือที่จำนวนมากกว่านี้ โดยเลขทุกจำนวนมีโอกาสปรากฏด้วยความน่าจะเป็นเท่ากัน
เช่นต้องการสุ่มหน่วยความจากบัญชีรายชื่อซึ่งได้จากสมาคมหน่วยความ สมมติมีสมาชิกอยู่
950 คน รายชื่อ 950 คนนี้เป็นกรอบตัวอย่าง เราจะให้เลขที่ 001 จนถึง 950 เป็นเลข
3 หลัก แล้วเราจะเลือกตัวเลข 3 หลัก จากตารางเลขสุ่มคือตาราง B-8 สมมติเราเริ่ม
ต้นจาก บรรทัดที่ 101 เอาเลขจากมุมซ้าย 3 ตัว แล้วต่อมาบรรทัดที่ 2 มุมซ้ายอีก 3 ตัว
และบรรทัดต่อไปเรื่อย ๆ ถ้าเลขใดซ้ำให้ตัดทิ้งไปจนครบตามต้องการเลขที่ได้แทนหมาย-
เลขของหน่วยความในบัญชีรายชื่อที่ใช้เป็นกรอบตัวอย่าง ดังนั้นหน่วยความที่ตกเป็นตัวอย่าง
คือ หมายเลข 132, 212, 001, 605, 912 เรื่อย ๆ ไปจนครบจำนวนตามต้องการ

การใช้ตารางเลขสัมพันธ์ตัวอย่าง

บรรทัด	(1) - (5)	หน่วยความ
101	13284	132
102	21224	212
103	99052	990 → ตัดทิ้งไปเพราะในรายชื่อมี คน
104	00199	001 เพียง 950 คน
105	60578	605
106	91240	912
107	97458	974 → ตัดทิ้งไปเพราะในรายชื่อมี
108	35249	352 เพียง 950 คน
109	38980	389
110	10750	107

Table B-8 Table of random digits

Line	(1)-(5)	(6)-(10)	(11)-(15)	(16)-(20)	(21)-(25)	(26)-(30)	(31)-(35)
101	13284	16834	74151	92027	24670	36665	00770
102	21224	00370	30420	03883	95648	89428	41583
103	99052	47887	81085	64933	66279	80432	65793
104	00199	50993	98603	38452	87890	94624	69721
105	60578	06483	28733	37867	07936	98710	98539
106	91240	18312	17441	01929	18163	69201	31211
107	97458	14229	12063	59611	32249	90466	33216
108	35249	38646	34475	72417	60514	69257	12480
109	38980	46600	11759	11900	46743	27860	77940
110	10750	52745	38749	87365	58959	53731	89295
111	36247	27850	73958	20673	37800	63835	71051
112	70994	66986	99744	72438	01174	42159	11392
113	99638	94702	11463	18148	81386	80431	90628
114	72055	15774	43857	99805	10419	76939	25993
115	24038	65541	85788	55835	38835	59399	13790
116	74976	14631	35908	28221	39470	91548	12854
117	35553	71628	70189	26436	63407	91178	90348
118	35676	12797	51434	82976	42010	26344	92920
119	74815	67532	72985	23183	02446	63594	98924
120	45246	88048	65173	50989	91060	89894	36036

ข้อสังเกต

1. มีหลายวิธีสำหรับเลือกตัวเลขจากตารางเลขสุ่ม โดยมีหลักการว่าต้องกำหนดหลักการล่วงหน้าว่าจะเริ่มต้นตรงไหน และการหยิบตัวเลขต้องทำแบบมีระบบเดียวกัน เช่น ต้องการเลข 3 ตัว จะอ่านตามบรรทัดจากซ้ายไปขวาจนจบบรรทัดแล้วขึ้นบรรทัดใหม่เหมือนอ่านหนังสือ แล้วตัดตัวเลขที่ละ 3 ตัว ไปเรื่อย ๆ หรือจะอ่านลงมาในคอลัมน์เดียวกัน แบบใดก็ได้ถ้าทำแบบเดียวกันตลอด
2. บางครั้งสมาชิกหรือหน่วยของประชากรมีเลขที่อยู่แล้ว เช่น ใบเสร็จรับเงิน ใบรับ-ส่งสินค้า รหัสนักศึกษา เราใช้ตัวเลขเหล่านั้นแทนหน่วยตัวอย่างได้เลย ถ้าเลขเหล่านั้นไม่มีการซ้ำซ้อนกัน
3. บางครั้งอาจไม่ต้องให้เลขที่กับรายชื่อในรอบตัวอย่าง เช่น ต้องการสุ่มพนักงานตัวอย่าง และมีประวัติการทำงานเก็บเข้าแฟ้มไว้อย่างดี สมมติมีพนักงาน 8,500 คน แฟ้มหนึ่งเก็บบันทึก 100 คน จึงต้องใช้ 85 แฟ้ม ถ้าได้เลขสุ่ม 0117 หมายความว่าอยู่ในแฟ้มที่ 1 และคนที่ 17 ถ้าได้เลข 0417 แสดงว่าต้องอยู่ในแฟ้มที่ 4 คนที่ 17 เป็นต้น

การสุ่มตัวอย่างแบบง่ายจากประชากรแบบไม่จำกัดจำนวน

(Simple Random Sampling From Infinite Population)

เมื่อประชากรเป็นแบบไม่จำกัดจำนวน ก็มีวิธีเดียวที่จะทราบข่าวสารจากประชากร คือจากตัวอย่าง เช่นสินค้าที่ผลิตจากเครื่องจักรมีจำนวนมากนับไม่ถ้วน เราไม่ทราบว่า N เป็นเท่าใด เพราะเครื่องทำการผลิตทุกวัน ดังนั้น สินค้า (ผลิตภัณฑ์) จากกระบวนการผลิตก็คือ ตัวอย่างสุ่มนั่นเอง

ข้อสังเกต

บางครั้งตัวอย่างจากประชากรแบบไม่จำกัด คือ ประชากร เช่นชิ้นส่วนอิเล็กทรอนิกส์สำหรับเครื่องคอมพิวเตอร์ที่ผลิตได้ใน 1 สัปดาห์ จะแทนตัวอย่างจากประชากรแบบไม่จำกัดซึ่งเกี่ยวกับกระบวนการผลิต แต่โรงงานส่งชิ้นส่วนนี้ไปให้ผู้ผลิตเครื่องคอมพิวเตอร์ ชิ้นส่วนเหล่านี้จะกลายเป็นประชากรของผู้ผลิตคอมพิวเตอร์ และผู้ผลิตคอมพิวเตอร์จะต้องมีแผนตรวจรับสินค้า โดยการสุ่มตัวอย่างจำนวนหนึ่งมาตรวจสอบคุณภาพ

ค่าสถิติที่ได้จากตัวอย่าง (Sample Statistics)

เมื่อเลือกตัวอย่างแบบสุ่มมา n จำนวนแล้ว เราจะวัดหรือหาค่าลักษณะที่สนใจ และแทนตัวแปร X_1, X_2, \dots, X_n จากข้อมูลนี้เราจะได้ค่าสถิติหลายอย่าง นั่นคือ เราเรียกค่าที่วัดจากตัวอย่างว่าค่าสถิติตัวอย่าง หรือเรียกสั้น ๆ ว่าค่าสถิติ เพื่อให้แตกต่างกับค่าที่วัดจากประชากรซึ่งเราเรียกว่า ค่าพารามิเตอร์ ค่าสถิติเบื้องต้นที่เราได้จากตัวอย่าง คือ

ค่าเฉลี่ยจากตัวอย่าง :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

ความแปรปรวนของตัวอย่าง :

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

และ

ค่าเบี่ยงเบนมาตรฐาน :

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{S^2}$$

นั่นคือ เราใช้ข่าวสารจากตัวอย่าง ประมาณค่าของประชากร ดังนี้

ประมาณค่า μ ด้วย \bar{X}

ประมาณค่า σ^2 ด้วย S^2

ประมาณค่า σ ด้วย S

และช่วงเชื่อมั่นของ μ คือ

$$\bar{X} \pm Z_{\alpha/2} S/\sqrt{n}$$

ในบทนี้เราจะพูดเฉพาะกรณีที่ $n \geq 30$ และ N เป็นค่าโตมาก แต่เป็นประชากรแบบจำกัด

และ n เป็นค่าเล็กมากเมื่อเทียบกับ N ดังนั้น สัดส่วนของตัวอย่างสุ่ม คือ n/N จึงเป็นค่าเล็ก และไม่ควรรุ่งสูงกว่า 5% มิฉะนั้นจะต้องปรับค่าด้วย finite population correction factor [$fpc = (N-n)/(N-1)$] และเป็นการสุ่มตัวอย่างแบบไม่มีการแทนที่ เพราะในทางปฏิบัติ การสุ่มแบบนี้เป็นที่นิยม เพราะเปิดโอกาสไม่ให้หน่วยที่ซ้ำกันตกเป็นตัวอย่างซึ่งจะทำให้ได้ข่าวสารซ้ำกับค่าเดิม จึงไม่ได้ประโยชน์เท่าที่ควร

นอกจากนี้ยังมีพารามิเตอร์อีกตัวหนึ่งที่น่าสนใจ คือ π ซึ่งคือสัดส่วนของลักษณะที่อยู่ในความสนใจซึ่งปรากฏในประชากร

$$\pi = \frac{y_1 + y_2 \dots + y_N}{N}, \quad \begin{array}{l} y = 0 \text{ ถ้าเป็นลักษณะที่ไม่สนใจ (F)} \\ y = 1 \text{ ถ้าเป็นลักษณะที่สนใจ (S)} \end{array}$$

และค่าประมาณของ π คือ

$$p = \frac{y_1 + y_2 \dots + y_n}{n}, \quad \begin{array}{l} y = 0 \text{ ถ้าเกิดลักษณะที่ไม่สนใจ (F)} \\ y = 1 \text{ ถ้าเกิดลักษณะที่สนใจ (S)} \end{array}$$

และช่วงเชื่อมั่นของ π คือ

$$p \pm Z_{\alpha/2} \sqrt{pq/n}$$

การสุ่มตัวอย่างแบบมีระบบ (Systematic random sampling)

การสุ่มตัวอย่างแบบมีระบบเป็นลักษณะหนึ่งของ convenient sample design มีวิธีการดังนี้

1. ต้องให้เลขที่กับหน่วยตัวอย่างในกรอบตัวอย่าง จาก 1, 2, ..., N
2. ถ้าต้องการสุ่มตัวอย่างขนาด n ต้องหาค่า sampling interval = $k = N/n$
3. หาหน่วยเริ่มต้นแบบสุ่ม เรียกว่า random start หรือ R โดยที่ $R \leq k$
4. ดังนั้น หน่วยแรกที่ใช้เป็นตัวอย่างคือ R

หน่วยที่ 2 คือ	คือ	$R + k$
หน่วยที่ 3 คือ	คือ	$R + 2k$
:		:
:		:

หน่วยสุดท้ายคือ n คือ $R + (n-1)k$
 ด้วยวิธีนี้จะได้อะไรจากบัญชีรายชื่อเป็นหน่วยที่ห่างกันด้วยระยะ k หน่วย
 ตลอด จำนวน n หน่วย

ผลของการสุ่มแบบนี้ ไม่ใช่ตัวอย่างแบบสุ่ม เนื่องจากหน่วยต่างๆ ที่ได้รับเลือกเป็นตัวอย่าง
 ไม่เป็นอิสระกัน และมีเพียงหน่วยตัวอย่างในกลุ่มแรกคือหน่วยที่ 1 ถึง k เท่านั้น ที่มีโอกาส
 ถูกเลือกแบบสุ่ม ตอนหาค่า R แต่เมื่อ R ถูกกำหนดแล้ว ตัวอย่างที่เหลือจะถูกกำหนดทันที
 ว่า จะต้องห่างกันด้วยช่วง k หน่วย

ถ้าหน่วยตัวอย่างในกรอบตัวอย่างมีลักษณะแบบเรียงลำดับ หรือมีลักษณะเป็นวัฏ-
 จักร ไม่ควรใช้การสุ่มตัวอย่างแบบมีระบบ เพราะอาจได้อะไรลักษณะเดียวซ้ำ ๆ กัน
 หมด เมื่อ k ไปพ้องกับรอบของวัฏจักร เช่น การเก็บบัตรประวัติพนักงาน ถ้าใบที่ 10,
 20, 30 แทนหัวหน้ากลุ่ม และถ้าเราได้ $k = 10$ ถ้า $R = 0$ หรือ 10, เราจะได้
 ตัวอย่างทั้งหมดเป็นหัวหน้ากลุ่ม ถ้า $1 \leq R \leq 9$ เราจะได้กลุ่มตัวอย่างที่ไม่มีหัวหน้า
 กลุ่มเลย ซึ่งตัวอย่างทั้ง 2 ชนิดนี้ ถือว่าไม่เป็นตัวแทนที่ดี

ข้อดีของการสุ่มแบบมีระบบ คือ ความสะดวก และควรใช้เมื่อแน่ใจว่าไม่มีลักษณะ
 การเปลี่ยนแปลงแบบมีระบบในประชากร (กรอบตัวอย่าง) ผลที่ได้จากการสุ่มแบบมีระบบ
 จะใกล้เคียงกับการสุ่มแบบง่าย จึงใช้สูตรคำนวณค่าประมาณต่าง ๆ ของการสุ่มแบบง่าย

แผนงานสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ

ถ้าเราสามารถจำแนกประชากรตามคุณลักษณะบางอย่าง เช่น เราจำแนกบุคคล
 ตามอายุ เพศ การศึกษา สถานภาพสมรส เป็นต้น เราอาจจำแนกธุรกิจตามขนาดของทุน
 และสินทรัพย์ จำนวนขาย จำนวนพนักงาน ลักษณะที่เราจำแนกนี้ เรียกว่า ชั้นภูมิ (stra-
 ta) และวิธีการแบ่งประชากรออกเป็นชั้นภูมิ เรียกว่า Stratification นั่นคือเรา
 สร้างชั้นภูมิจากคุณลักษณะที่เราทราบ การแบ่งชั้นภูมิของประชากรมีหลักการ 2 ข้อ คือ

1. ต้องให้ความผันแปรภายในชั้นภูมิเดียวกันมีค่าน้อยที่สุด นั่นคือต้องมีความ
 เป็นเอกภาพมากที่สุด
2. ควรให้ความผันแปรระหว่างชั้นภูมิต่าง ๆ มากที่สุด คือมีความแตกต่าง
 อย่างเห็นชัดระหว่างชั้นภูมิต่าง ๆ

ปัญหาที่ตามมา

1. จะใช้คุณลักษณะใดเป็นหลักในการแบ่งชั้นภูมิ?
2. จะแบ่งสรร n ตัวอย่าง ให้แต่ละชั้นภูมิเท่าใด?
3. เมื่อได้ข้อมูลจากชั้นภูมิต่าง ๆ แล้ว จะวิเคราะห์อย่างไร?

คำตอบคือ

1. คุณลักษณะที่สมควรเลือกใช้เป็นเกณฑ์แบ่งชั้นภูมิคือ คุณลักษณะที่เราทราบว่าทำให้ข้อมูลภายในชั้นภูมิเดียวกันมีความผันแปรน้อยที่สุด และข้อมูลจากชั้นภูมิที่ต่างกัน มีความผันแปรมากที่สุด เช่นการศึกษาความพอใจของพนักงานต่อระบบสวัสดิการต่าง ๆ เช่น เงินตอบแทน, บำนาญ, ค่ารักษาพยาบาล, การประกันชีวิต, การให้หยุดพักผ่อน ความเห็นของพนักงานอาจจะแตกต่างกันระหว่างเพศและอายุของพนักงาน ปกติจะแบ่งประชากรเป็นชั้นภูมิ ตามเพศ และอายุ ก่อนการสุ่มตัวอย่าง
2. จะจัดสรรขนาดตัวอย่างให้แต่ละชั้นภูมิอย่างไร? วิธีที่ควรพิจารณาใช้มี 2 วิธี คือ

2.1 การจัดสรรแบบสัดส่วน (Proportional allocation)

2.2 การจัดสรรแบบ Optimum (Optimum allocation)

การจัดสรรแบบสัดส่วน

สมมติ กรุงเทพฯ มีประชากรผู้ใหญ่ 5 ล้านคน เป็นชาย 2 ล้าน และหญิง 3 ล้าน ถ้าเราใช้หน่วยเป็นล้านและแบ่งประชากรตามเพศ จะได้ 2 ชั้นภูมิ คือ ชาย และหญิง

$$N_{\text{ชาย}} = 2, N_{\text{หญิง}} = 3$$

ถ้าจัดสรรแบบเป็นสัดส่วน นั่นคือต้องสุ่มในอัตราส่วน ช : ญ = 2 : 3 หรือ 40% กับ 60% เช่นต้องการตัวอย่าง $n = 400$, $n_{\text{ชาย}} = .40(400) = 160$ คน และ $n_{\text{หญิง}} = .60(400) = 240$ คน

การจัดสรรแบบ optimum

ถ้าชั้นภูมิต่าง ๆ มีความผันแปรแตกต่างกัน เช่น สมมติ $\sigma_{\text{ชาย}} = 406$ และ

$\sigma_{ny} = 300$ วิธีจัดสรรตัวอย่างแบบ optimum จะให้ความแม่นยำสูงสุด เพราะจะพิจารณาทั้งขนาด (N_n) และความผันแปร (σ_n^2) ดังนี้

$$\begin{aligned} n_{\text{ช}} : n_{\text{ญ}} &= N_{\text{ช}} : N_{\text{ญ}} \\ &= (2 \times 406) : (3 \times 300) \\ &= 812 : 900 \\ &= 812/1712 : 900/1712 \\ &= .474 : .526 \quad \text{หรือ} \quad 47.4\% : 52.6\% \end{aligned}$$

ดังนั้น ถ้า $n = 400$ คน จะจัดสรรให้ $n_{\text{ญ}} = .474(400) = 190$ คน, $n_{\text{ช}} = .526(400) = 210$ หากค่าใช้จ่ายในการสัมภาษณ์ต่อหน่วยตัวอย่างไม่เท่ากัน บางชั้นภูมิค่า บางชั้นภูมิสูง เช่น สมมติค่าใช้จ่ายในการสัมภาษณ์ เพศชาย หน่วยละ 16 บาท เพศหญิง หน่วยละ 9 บาท และมีงบประมาณอยู่ทั้งสิ้น 4,000 บาท เช่นนี้ จะต้องนำค่าใช้จ่ายต่อหน่วยเข้ามาร่วมพิจารณาด้วย โดยมีหลักว่า ชั้นภูมิที่เสียค่าใช้จ่ายต่อหน่วยสูง (ต้องเดินทางไกล ฯลฯ) จะใช้ขนาดตัวอย่างน้อยกว่าชั้นภูมิที่มีค่าใช้จ่ายต่อหน่วยต่ำ นั่นคือ ขนาดตัวอย่างของชั้นภูมิจะเป็นสัดส่วนแบบผกผันกับค่าใช้จ่ายต่อหน่วย นั่นคือ

$$\begin{aligned} n_{\text{ช}} : n_{\text{ญ}} &= N_{\text{ช}} \sigma_{\text{ช}} / \sqrt{C_{\text{ช}}} : N_{\text{ญ}} \sigma_{\text{ญ}} / \sqrt{C_{\text{ญ}}} \\ n_{\text{ช}} : n_{\text{ญ}} &= 2(406) / \sqrt{16} : 3(300) / \sqrt{9} \\ &= 203 : 300 \quad \text{หรือ} \quad 203/503 : 300/503 \\ &= .4036 : .5964 \quad \text{หรือ} \quad 40.36\% : 59.64\% \end{aligned}$$

แต่มีงบประมาณอยู่ 4,000 บาท จึงต้องหาขนาดตัวอย่างรวม (n) ให้สอดคล้องกับเงิน 4,000 บาทก่อน ดังนี้

$$\begin{aligned} .4036(n)(16) + .5964(n)(9) &= 4000 \\ 6.4576n + 5.3676n &= 4000 \\ 11.8252n &= 4000 \\ n &= 338.26 = 338 \quad \text{คน} \end{aligned}$$

และจัดสรรให้ 2 กลุ่ม ดังนี้

$$\begin{aligned} n_{\text{ช}} &= .4036(338) = 136.42 = 136 \quad \text{คน} \\ n_{\text{ญ}} &= .5964(338) = 201.58 = 202 \quad \text{คน} \end{aligned}$$

ในทางปฏิบัติ เรามักไม่ทราบค่า σ_h , σ_{μ} และถ้าไม่สามารถหาค่าประมาณที่เหมาะสมได้ ก็ควรใช้การจัดสรรแบบสัดส่วน

โดยทั่ว ๆ ไป การสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ มิได้มีเพียง 2 ชั้นภูมิจะมีทั้งหมด L ชั้นภูมิ และต้องจัดสรร n ตัวอย่าง สำหรับ L ชั้นภูมิ ให้ n_h แทนขนาดตัวอย่างของชั้นภูมิที่ h และ

$$n_1 + n_2 + \dots + n_h = n$$

ถ้าใช้ optimum allocation และคำนึงถึงค่าใช้จ่ายต่อหน่วยคือ C_h

$$n_h = \frac{N_h \sigma_h / \sqrt{C_h}}{\sum N_h \sigma_h / \sqrt{C_h}} \times (n)$$

ถ้า C_h ไม่ต่างกันระหว่างชั้นภูมิ

$$n_h = \frac{N_h \sigma_h}{\sum N_h \sigma_h} \times (n)$$

และถ้า σ_h ของแต่ละชั้นภูมิ ไม่ต่างกันมาก

$$n_h = \frac{N_h}{N} \times (n)$$

แผนการสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage Sampling)

การสุ่มแบบง่ายและแบบแบ่งชั้นภูมิ ในบางกรณีอาจให้หน่วยตัวอย่างกระจุกจัดกระจายเป็นบริเวณกว้าง ทำให้เสียเวลาและค่าใช้จ่ายสูง เช่นการสัมภาษณ์ผู้มีสิทธิออกเสียงเลือกตั้ง จะมีหน่วยตัวอย่างกระจายอยู่ทั่วทั้งประเทศ ผู้สำรวจ (ผู้สัมภาษณ์) จะต้องเสียเวลาและค่าใช้จ่ายในการเดินทางจากจุดหนึ่ง ไปยังอีกจุดหนึ่ง ซึ่งเวลาที่เสียไปสำหรับการเดินทางนั้นจะมากกว่าเวลาที่ใช้ในการสัมภาษณ์เสียอีก ดังนั้นจึงเป็นการดีถ้าสามารถจัดให้หน่วยตัวอย่างทั้งหลายอยู่ใกล้ ๆ กัน เช่นแทนที่จะสัมภาษณ์ 1350 ครัวเรือนที่กระจายอยู่ทั่วประเทศถ้าผู้สุ่มใช้วิธีการสุ่มแบบง่ายหรือแบบชั้นภูมิ เราอาจเสียค่าใช้จ่ายจังหวัดตัวอย่างมา 10 จังหวัด จากแต่ละจังหวัดตัวอย่างสุ่มมาจังหวัดละ 3 อำเภอตัว-

อย่าง แต่ละอำเภอตัวอย่าง สุ่มมา 3 ตำบลตัวอย่าง แต่ละตำบลตัวอย่างสุ่มมา 3 หมู่บ้านตัวอย่าง แต่ละหมู่บ้านตัวอย่างสุ่มมา 5 ครัวเรือนตัวอย่าง จะได้ $10 \times 3 \times 3 \times 3 \times 5 = 1350$ ครัวเรือนตัวอย่าง การสุ่มวิธีนี้ข้อดีคือ ไม่ต้องทำ frame ของ ครัวเรือนทั่วประเทศ เป็นการประหยัดเวลาและค่าใช้จ่าย และครัวเรือนที่ตกเป็นตัวอย่าง จะไม่กระจายทั่วไป เป็นการลดค่าใช้จ่ายในการเดินทาง

วิธีการสุ่มแบบนี้เรียกว่า การสุ่มแบบหลายขั้นตอน จากตัวอย่างมีทั้งหมด 5 ขั้นตอน ถ้ามีเพียง 2 ขั้นตอนเรียกว่า two-stage sampling หรือการสุ่มแบบสองขั้นตอน เช่น ขั้นที่หนึ่ง สุ่มมา 10 จังหวัดตัวอย่าง ขั้นที่สอง สุ่มมาจังหวัดละ 3 อำเภอ จะได้ทั้งหมด $10 \times 3 = 30$ อำเภอตัวอย่าง แล้วสัมภาษณ์ครัวเรือนทุกครัวเรือนใน 30 อำเภอตัวอย่างนั้น แต่ถ้าทำเพียงขั้นตอนเดียว เรียกว่า single-stage sampling หรือการสุ่มแบบขั้นเดียว คือพอได้ 10 จังหวัดตัวอย่างแล้ว ต้องสัมภาษณ์ทุกครัวเรือนใน 10 จังหวัดตัวอย่างนั้น จะเห็นว่าจะได้ตัวอย่างมากเกินไป และเสียเวลาและค่าใช้จ่ายมากโดยไม่จำเป็น จึงนิยมแบบหลายขั้นตอน สำหรับหน่วยตัวอย่างในแต่ละขั้นตอน จะมีชื่อเรียกเฉพาะ เช่น จากตัวอย่าง หน่วยตัวอย่างในขั้นตอนแรกคือ จังหวัด ทำหน้าที่เป็นหน่วยตัวอย่างปฐมภูมิ หรือ primary sampling unit เรียกย่อ ๆ ว่า psu อำเภอ ทำหน้าที่เป็นหน่วยตัวอย่างในขั้นตอนที่สอง เรียกว่าหน่วยตัวอย่างทุติยภูมิ หรือ secondary sampling unit ตำบล ทำหน้าที่เป็นหน่วยตัวอย่างตติยภูมิ หมู่บ้าน เป็นหน่วยตัวอย่างในขั้นตอนที่สี่ และ ครัวเรือน เป็นหน่วยตัวอย่างในขั้นตอนสุดท้าย เรียกว่า elementary sampling unit ข้อสังเกตคือ ในแต่ละขั้นตอนของการสุ่มแบบหลายขั้นตอน นิยมใช้วิธีการสุ่มแบบง่ายในการเลือกหน่วยตัวอย่าง และถ้าหน่วยตัวอย่างเป็นท้องที่ จะเรียกว่า Area Sampling ซึ่งสะดวกในทางปฏิบัติเพราะมีแผนที่แสดงการแบ่งภูมิภาค อนึ่งการสุ่มแบบหลายขั้นตอนนี้ บางครั้งอาจเรียก การสุ่มแบบกลุ่ม หรือ Cluster Sampling เพราะในแต่ละขั้นตอนที่สุ่มตัวอย่างจะได้หน่วยตัวอย่างในลักษณะเป็นกลุ่ม ซึ่งถือเป็นข้อได้เปรียบของวิธีสุ่มแบบนี้ในเชิงเศรษฐศาสตร์ เพราะช่วยประหยัดค่าใช้จ่ายในการสร้าง frame ของ elementary sampling unit (ถ้าใช้การสุ่มแบบง่าย) การสร้าง frame แบบกลุ่มจะง่ายกว่า เพราะสามารถหารายชื่อที่สมบูรณ์ได้จากระบบทะเบียน (รายชื่อจังหวัด อำเภอ ตำบล หมู่บ้าน)

แบบฝึกหัด

1. จงเปรียบเทียบข้อดี-ข้อเสีย ระหว่างวิธีการสุ่มแบบสุ่ม และวิธีการสำรวจ
2. จงอธิบายว่าเหตุใดการเลือกตัวอย่างแบบสุ่ม เป็นสิ่งจำเป็นสำหรับแผนแบบการสุ่มตัวอย่างที่ดี
3. การแบ่งประชากรเป็นชั้นภูมิมีผลกระทบต่อการศึกษาเลือกเป็นตัวอย่างแบบสุ่มอย่างไร
4. จงอธิบายข้อได้เปรียบของแผนการสุ่มตัวอย่าง โดยแบ่งประชากรเป็นชั้นภูมิ
5. จงอธิบายความหมายของแผนการสุ่มตัวอย่างแบบหลายขั้นตอน
6. การสุ่มตัวอย่างแบบโควตาต่างกับการเลือกตัวอย่างแบบสุ่มอย่างไร
7. สมมติว่าสุ่มผู้มีอายุเกิน 18 ปี ในจังหวัดกรุงเทพฯ มา 400 คน คำถามหนึ่งที่สัมพันธ์คือ "ท่านทำงานสัปดาห์ละ 35 ชั่วโมงขึ้นไป ในปีที่แล้วเป็นจำนวนกี่สัปดาห์?" สมมติได้ $\bar{X} = 30$, $S = 10.5$, ให้ประชากรผู้ใหญ่จังหวัดกรุงเทพฯ มี 5 ล้านคน
 - ก) จงประมาณจำนวนสัปดาห์ที่เฉลี่ยวสำหรับประชากรทั้งจังหวัด และสร้างช่วงเชื่อมั่น 95%
 - ข) จงประมาณจำนวนรวมของสัปดาห์ที่ประชากรจังหวัดกรุงเทพฯ ทำงานเกินสัปดาห์ละ 35 ชั่วโมง และสร้างช่วงเชื่อมั่น 95%
 - ค) จงอธิบายความหมายของผลที่ได้ในข้อ (ก) และ (ข)
8. จากข้อ 7. ถ้าประชากรผู้ใหญ่ในจังหวัดกรุงเทพฯ มี 5 ล้านคน และจากตัวอย่างที่สุ่มมา 400 คน เป็นชาย 170 คน และหญิง 230 คน (สุ่มแบบง่าย) และมีข้อมูลสรุป ดังนี้

	\bar{X}	S
ชาย	40	2.7
หญิง	22.6	7.6

 - ก) จงสร้างช่วงเชื่อมั่น 95% ของจำนวนสัปดาห์การทำงานโดยเฉลี่ยของประชากรชายในกรุงเทพฯ ในปีที่แล้ว
 - ข) ทำเหมือนข้อ (ก) แต่หาช่วงเชื่อมั่นของประชากรหญิง (ที่มีงานทำ)
 - ค) เปรียบเทียบช่วงเชื่อมั่นระหว่างข้อ (ก) และ (ข) กับข้อ 7(ก) เหตุใดช่วงเชื่อมั่นจึงมีขนาดที่แตกต่างกัน
9. ในการศึกษาการทำงานของพนักงานผู้หนึ่ง ถ้าใช้วิธีสังเกตการทำงานของเขาในช่วงเวลาที่ห่างเท่าๆ กัน เช่นทุก 15 นาที โดยใช้การถ่ายภาพ หรือ TV หรือให้คนสังเกตการณ์ วิธีนี้ คือการสุ่มตัวอย่างแบบมีระบบ ท่านคิดว่าเหมาะสมหรือไม่

10. ถ้าสินค้าที่พนักงานผู้นั้นผลิต จะได้รับการบรรจุใส่หีบห่อเก็บไว้ในห้องเก็บสินค้า และมีการตรวจสอบสภาพสินค้าคงคลังเป็นประจำทุกสัปดาห์ เพื่อคัดสินค้าที่เน่าเสียทิ้งไป ท่านจะใช้วิธีคัดเลือกแต่ละหน่วยด้วยการสุ่มแบบมีระบบไหม? เช่นการตรวจทุก ๆ กล้องที่ 50 โดยนับจากซ้ายไปขวา หรือจากข้างล่างขึ้นข้างบน หรือจากด้านหน้าไปด้านหลัง ของห้องเก็บสินค้าคงคลัง? เพราะเหตุใด?
11. บริษัทหนึ่งต้องการสุ่มพนักงานตัวอย่างมาจำนวน 500 คน และมั่นใจว่าตัวแปรที่มีอิทธิพลต่อข้อมูลที่เก็บมาคืออายุการทำงานกับบริษัท ดังนั้นจึงจะแบ่งเป็นชั้นภูมิตามอายุการทำงาน ซึ่งบริษัทมีข้อมูลดังนี้

อายุการทำงาน	h	N_h	σ_h
น้อยกว่า 2 ปี	1	2,000	.7
2 - 5 ปี	2	1,000	1.4
5 ปีขึ้นไป	3	1,000	2.8

- ก) จงหาขนาดตัวอย่าง n_h ของแต่ละชั้นภูมิ โดยวิธีจัดสรรแบบเป็นสัดส่วน
- ข) จงจัดสรรแบบ optimum
- ค) เหตุใดผลที่ได้ในข้อ (ก) และ (ข) จึงต่างกัน
12. จากข้อ 11. ถ้าได้ข้อมูลจากตัวอย่าง คือ X = มูลค่าการถือหุ้นเป็นบาท ดังนี้

อายุการทำงาน	h	\bar{X}_h	S_h
น้อยกว่า 2 ปี	1	60	25
2 - 5 ปี	2	200	60
มากกว่า 5 ปี	3	2,500	300

- ก) สมมติว่าข้อมูลมาจากการจัดสรรแบบเป็นสัดส่วนกับขนาดของประชากร จงหาอายุการทำงานเฉลี่ยของพนักงานทั้งหมด

- ข) สมมติว่าเป็นการจัดสรรแบบ optimum จงหาอายุการทำงานเฉลี่ยของพนักงานทั้งหมด
- ค) จงเปรียบเทียบค่าเฉลี่ยจาก (ก) และ (ข) และวิจารณ์
13. จากข้อ 11. สมมติได้ข้อมูลจากตัวอย่าง คือ $X =$ จำนวนปีที่สำเร็จการศึกษา ดังนี้

อายุการทำงาน	h	\bar{X}_h	S_h
ต่ำกว่า 2 ปี	1	13	3
2 - 5 ปี	2	12	2.5
เกิน 5 ปี	3	10	2

จงประมาณจำนวนปีที่จบการศึกษาโดยถัวเฉลี่ย โดยวิธีจัดสรรแบบสัดส่วน และวิธี optimum

14. งานสำรวจชิ้นหนึ่ง ต้องทำการสำรวจจากครอบครัวตัวอย่างซึ่งมีรายได้ต่ำ ซึ่งอาศัยอยู่ในชุมชนแออัด ครอบครัวในท้องที่ตัวอย่างแบ่งเป็น 2 ประเภท คือ มีการศึกษาและไม่มีการศึกษา (ยึดหัวหน้าครอบครัวเป็นหลัก) สำหรับครอบครัวที่ไม่มีการศึกษา ต้องใช้พนักงานสำรวจที่มีประสิทธิภาพสูง เพื่อการสื่อความหมายที่ดีและได้ข้อมูลที่เชื่อถือได้ จึงต้องใช้พนักงานที่มีความรู้พิเศษ และจะต้องเสียค่าใช้จ่ายสูงกว่าปกติ ข้อมูลของการสำรวจมีดังนี้

การศึกษา	จำนวนครัวเรือน	ค่าใช้จ่ายต่อหน่วย	S_h
มีการศึกษา	6,000	9	500
ไม่มีการศึกษา	4,000	16	800

- ก) จงจัดสรรตัวอย่างขนาด 165 ครัวเรือนให้กับชั้นภูมิทั้ง 2 โดยวิธีสัดส่วน
- ข) จงจัดสรรโดยวิธี optimum
- ค) ถ้ามีงบประมาณสำหรับสัมภาษณ์ 2,000 บาท จงจัดสรรให้แก่ 2 ชั้นภูมิ

ง) จะต้องเสียค่าใช้จ่ายในการสัมภาษณ์ แต่ละชั้นภูมิเป็นเงินเท่าใด?

15. คลังเก็บสินค้าของตัวแทนจำหน่ายผลิตภัณฑ์ไฟฟ้าเสียหายเล็กน้อย เนื่องจากพายุไต้ฝุ่น กิจการต้องการเปิดกล่องสินค้าเพื่อสำรวจความเสียหายเนื่องจากโดนน้ำจำนวนหนึ่ง ผลความเสียหายส่วนใหญ่เกิดกับชั้นล่างของคลังเก็บสินค้า ดังนั้นจึงจะแบ่งโดยใช้ชั้นต่าง ๆ เป็นชั้นภูมิ ผลการสำรวจ ได้ข้อมูล ดังนี้

ชั้น	จำนวนรวม	จำนวนกล่องที่เปิดตรวจ	จำนวนกล่องที่เสีย
1	10,000	60	30
2	12,000	50	15
3	8,000	40	8

จงประมาณเปอร์เซ็นต์ของเครื่องไฟฟ้า ที่เสียหายเพราะน้ำท่วม

16. ฝ่ายบริหารบุคคลได้เลือกโรงงานในเครือมา 5 โรง เพื่อทดลองใช้ระบบสวัสดิการแบบใหม่ โดยพิจารณาว่าโรงงาน 5 โรงนั้นจะสามารถใช้เป็นตัวแทนของสภาวะต่าง ๆ ของโรงงานทั้งหมดในเครือ โรงงาน 5 โรงนี้เป็นตัวอย่างแบบใด
17. ถ้าพรุ่งนี้จะมีการสุ่มประชากรที่บรรลุนิติภาวะจากเมือง ๆ หนึ่ง ซึ่งอาจมีบางคนไปถูกระนอกตัวเมือง จงหาประชากรเป้าหมาย และประชากรตัวอย่าง

เฉลยแบบฝึกหัด

7. $n = 400, \bar{x} = 30, S = 10.5, N = 5,000,000$

(ก) $s_{\bar{x}} = S/\sqrt{n} = 10.5/\sqrt{400} = 1.67$

ช่วงเชื่อมั่น 95% ของ μ คือ $\bar{x} \pm Z_{.025} S_{\bar{x}}$

$$= 30 \pm (1.96)(1.67)$$

$$= 30 \pm 3.25 = 26.7464, 33.2536 \text{ สัปดาห์}$$

(ข) $\hat{T} = N\bar{x} = 5,000,000(30) = 150,000,000 = 150 \text{ ล้านสัปดาห์}$

ช่วงเชื่อมั่น 95% ของ T คือ $\hat{T} \pm N(Z_{\alpha/2} S_{\bar{x}})$

$$= 150 \pm 5(1.96)(1.67)$$

$$= 150 \pm 16.268 = 133.732, 166.268 \text{ ล้านสัปดาห์}$$

(ค) ความแตกต่างระหว่างข้อ (ก) และ (ข) คือ ข้อ (ก) เป็นการสรุปด้วยความเชื่อมั่น 95% ว่า ประชาชนในกรุงเทพมหานครทำงานเกิน 35 ชั่วโมงต่อสัปดาห์ ในปีที่แล้ว โดยเฉลี่ยคนละ 26.7464 - 33.2536 สัปดาห์ หรือโดยประมาณ 27 - 33 สัปดาห์ ส่วนในข้อ (ข) เป็นจำนวนรวมของประชากรทั้ง 5 ล้านคน ในปีก่อนที่ทำงานเกินสัปดาห์ละ 35 ชั่วโมง มีอยู่ระหว่าง 133.732 - 166.268 ล้านสัปดาห์

8. $N = 5 \text{ ล้านคน}, n = 400, n_{\text{ช}} = 170, n_{\text{ญ}} = 230$

$\bar{x}_1 = 40, \bar{x}_2 = 22.6, S_1 = 2.7, S_2 = 7.6$

(ก) 35% ช่วงเชื่อมั่นของ $\mu_{\text{ช}}$ = $\bar{x}_1 \pm Z_{.025} S_{\bar{x}_1}$

$$= 40 \pm (1.96)(2.7\sqrt{1700})$$

$S_{\bar{x}_1} = 207.207$	= 40 ± .4058775
---------------------------	-----------------

$$= 39.594123, 40.405877$$

(ข) 95% ช่วงเชื่อมั่นของ $\mu_{\text{ญ}}$ = $\bar{x}_2 \pm Z_{.025} S_{\bar{x}_2}$

$$= 22.6 \pm (1.96)(7.6/\sqrt{230})$$

$$= 22.6 \pm (1.96)(.5011291)$$

$$= 22.6 \pm 0.982213$$

$$= 21.617787, 23.582213$$

(ค) จากข้อ 7(ก) มี $S_{\bar{x}} = 1.67$ ซึ่งโตกว่า $S_{\bar{x}_1}$ และ $S_{\bar{x}_2}$ จึงทำให้ช่วงเชื่อมั่นกว้างกว่า

9. ไม่เหมาะสม เพราะพนักงานอาจรู้ตัวว่ามีการตรวจในระยะห่างสม่ำเสมอ เขาอาจ "พราง" ได้ ถ้าเขาสามารถกะประมาณช่วงการตรวจได้
10. ขึ้นอยู่กับวิธีการจัดเก็บสินค้า ถ้าทุกหน่วยที่ 50 อยู่ชั้นล่างเสมอ หรืออยู่ทางด้านหลังสุดเสมอ อาจเกิดการได้เปรียบ หรือเสียเปรียบจากสิ่งแวดล้อม เช่น อาจมีปัจจัยที่ทำให้ของชำรุดง่าย หรือลดอัตราชำรุด แต่ถ้าวัดเก็บเป็นแบบสุ่ม การใช้วิธีสุ่มแบบง่ายและแบบมีระบบจะให้ผลไม่ต่างกัน
11. บริษัทหนึ่งต้องการส่งพนักงานตัวอย่าง 500 คน โดยมี

$$N_1 = 2000, N_2 = 1000, N_3 = 1000, N = 4000, \sigma_1 = .7, \sigma_2 = 1.4, \sigma_3 = 2.8$$

(ก) วิธีจัดสรรแบบสัดส่วน

$$n_h = \frac{N_h}{N} \times n$$

$$n_1 = (2000/4000) \times 500 = 250 \text{ คน}$$

$$n_2 = (1000/4000) \times 500 = 125 \text{ คน} = n_3$$

(ข) จัดสรรวิธี optimum

$$n_h = \frac{N_h \sigma_h}{\sum N_h \sigma_h} \times n$$

ชั้นภูมิ	N_h	σ_h	$N_h \sigma_h$	$(N_h \sigma_h / \sum N_h \sigma_h) \times n$
1	2000	.7	1400	$(1400/5600) 500 = 125 = n_1$
2	1000	1.4	1400	$(1400/5600) 500 = 125 = n_2$
3	1000	2.8	2800	$(2800/5600) 500 = 250 = n_3$
	$N = 4000$		5600	500 = n

(ค) ผลต่างของ (ก) และ (ข) เพราะแบบ (ข) เอาความผันแปรมาพิจารณา กลุ่มที่มีความผันแปรสูง จะต้องใช้ขนาดตัวอย่างโต เช่น กลุ่มที่ 3 ถ้าความผันแปรของทุก ๆ กลุ่มเท่ากัน ผลจาก (ก) และ (ข) จะเหมือนกัน

$$12. \quad \bar{X} = \sum N_h \bar{x}_h / N$$

$$(ก) \text{ วิธีสัดส่วน} = \{2000(60) + 1000(20) + 1000(2500)\} / 4000 \\ = 705 \text{ บาท}$$

$$n_1 = 250, n_2 = n_3 = 125, n = 500$$

(ข) จัดสรรแบบ optimum นั่นคือ $n_1 = 125, n_2 = 125, n_3 = 250$

$$\bar{x} = 705 \text{ เท่าเดิม}$$

(ค) ค่า \bar{x} จาก 2 วิธี ได้เท่ากัน เพราะไม่ได้เอาขนาดตัวอย่าง n_h (ซึ่งอาจต่างกัน) มาถ่วงน้ำหนัก แต่ใช้ถ่วงน้ำหนักด้วยขนาดของชั้นภูมิ N_h ซึ่งมีค่าคงเดิมไม่ว่าจะใช้วิธีสัดส่วน หรือ optimum

$$3. \quad \bar{x}_1 = 13, \bar{x}_2 = 12, \bar{x}_3 = 10, N_1 = 2000, N_2 = 1000, N_3 = 1000 \\ S_1^2 = 9, S_2^2 = 6.25, S_3^2 = 4. (N_1/N) = .5, (N_2/N) = (N_3/N) \\ = .25$$

การจัดสรรแบบสัดส่วน 9 นั่นคือ $n_1 = 250, n_2 = n_3 = 125$

$$\bar{x} = \sum N_h \bar{x}_h / N \\ = \{2000(13) + 1000(12) + 1000(10)\} / 4000 \\ = 12$$

การจัดสรรแบบ optimum $n_1 = 125, n_2 = 125, n_3 = 250$

$$\bar{x} = 12 \text{ ปี เท่าวิธีจัดสรรแบบสัดส่วน}$$

$$14. \quad n = 165, N_1 = 6000, N_2 = 4000, N = 10,000$$

(ก) จัดสรรโดยวิธีสัดส่วน

$$n_h = (N_h/n) \times n$$

$$n_1 = (6,000/10,000) 165 = 99$$

$$n_2 = (4,000/10,000) 165 = 66$$

(ข) จัดสรรวิธี optimum

ชั้นภูมิ	N_h	S_h	$N_h S_h$	$N_h S_h / \sqrt{C_h}$	$\sqrt{C_h}$	$N_h S_h \sqrt{C_h}$
1	6000	500	3,000,000	1,000,000	3	9,000,000
2	4000	800	3,200,000	800,000	4	12,800,000
			6,200,000	1,800,000		21,800,000

$$n_1 = (3,000,000/6,200,000) 165 = 80$$

$$n_2 = (3,200,000/6,200,000) 165 = 85$$

(ค) มีงบประมาณ 2000 บาท ต้องการ n ก่อน

$$n = \frac{\sum N_h S_h / \sqrt{C_h}}{\sum N_h S_h \sqrt{C_h}} \times C$$

$$= (1,800,000/21,800,000) \times 2000 = 165$$

$$n_1 = (1,000,000/1,800,000) \times 165 = 92$$

$$n_2 = (800,000/1,800,000) \times 165 = 73$$

(ง) ชั้นภูมิที่ 1, $n_1 = 92$, $C_1 = 9$

$$\text{ค่าใช้จ่ายรวม} = 92 \times 9 = 828 \text{ บาท}$$

ชั้นภูมิที่ 2, $n_2 = 73$, $C_2 = 16$

$$\text{ค่าใช้จ่ายรวม} = 73 \times 16 = 1,168$$

$$\text{รวม 2 ชั้นภูมิ} = 828 + 1,168 = 1,996 \text{ บาท}$$

$$15. \quad p_1 = 30/60 = .5 \qquad N_1 = 10,000, \quad N_2 = 12,000$$

$$p_2 = 15/50 = .3 \qquad N_3 = 8,000, \quad N = 30,000$$

$$p_3 = 8/40 = .2$$

$$p = \{ \sum N_h p_h \} / N$$

$$= \{ 10,000(.5) + 12,000(.3) + 8,000(.2) \} / 30,000$$

$$= .34$$

16. เป็น judgement sample เพราะผู้บริหารได้พิจารณาแล้วว่าโรงงานทั้ง 5 นั้นมีลักษณะเป็น "ตัวแทน" ของโรงงานอื่น ๆ

17. ประชากรเป้าหมาย คือผู้บรรลุนิติภาวะทั้งหมดในเมืองนั้น
 ประชากรตัวอย่าง คือผู้บรรลุนิติภาวะที่ปฏิบัติงานในเมืองนั้นในขณะสุ่มตัวอย่าง