บทที่ 3 แผนสำรวจ

การสำมะโน

การสำรวจ

Nonprobability Sampling

Probability sampling

-การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling)

-การสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling)

-การสุ่มตัวอย่างแบบมีชั้นภูมิ (Stratified Sampling)

บทที่ 3 แผนสำรวจ

การเก็บข้อมูลจากประชากรที่อยู่ในขอบข่าย (Coverage) ของงานวิจัยนั้นเราสามารถจัดเก็บได้ 2 วิธีคือ

1. การสำมะโน (Census)
2. การสำรวจด้วยตัวอย่าง (Sample Survey)

1. การสำมะโน (Census) มีความหมายดังนี้คือ ผู้วิจัยจะทำการเก็บข้อมูลจากทุกหน่วย (Unit) ในประชากรโดยไม่มีการละเว้น ตัวอย่างของงานเช่นนี้คือ การทำสำมะโนประชากรของประเทศไทย ซึ่งจะทำการสำรวจทุกครัวเรือนในประเทศโดยที่จะทำ 10 ปีต่อครั้ง การเก็บข้อมูลโดยวิธีนี้มีข้อดีตรงที่สามารถศึกษาข้อมูลแยกรายละเอียดลงไปได้ตามหน่วยย่อยเช่นเขตพื้นที่เป็นต้นและได้ข้อมูลครบถ้วน แต่มีข้อเสียคือไม่สามารถจะทำได้ในเวลาอันรวดเร็วได้ จึงทำให้เสียค่าใช้จ่ายสูงและใช้เวลานาน ในกรณีต้องการผลรีบด่วนไม่สามารถจะทำได้

2. การสำรวจด้วยตัวอย่าง (Sample Survey) เป็นกระบวนการที่เราใช้ข้อมูลจากบางหน่วยซึ่งเป็นตัวแทนของประชากรเท่านั้น เพื่อนำข้อมูลที่ได้มาไปอนุมานเข้าสู่ประชากรอีกทีหนึ่ง วิธีนี้มีข้อดีที่เห็นได้ชัดคือ เสียเวลาและค่าใช้จ่ายน้อย เหมาะกับสถานการณ์ที่ต้องการผลอย่างรีบด่วนเพื่อนำไปปฏิบัติงานเช่น การศึกษาการแพร่ของโรคระบาด การศึกษาการเคลื่อนย้ายของประชากรและแม้กระทั่งในบางกรณีย์เช่นการทดสอบคุณภาพสินค้าซึ่งการทดสอบนั้นทำให้สินค้าเสียไปเลย เช่นการทดสอบนมที่บรรจุในกระป๋องว่ามีคุณภาพเป็นอย่างไร ผู้ศึกษาต้องทดสอบโดยการเจาะนมจากกระป๋องมาทดสอบ วิธีเช่นนี้สินค้านั้นจะเสียไปเลยเราจึงไม่สามารถจะใช้วิธีการสำมะโนในการเก็บข้อมูลได้จึงต้องพึ่งพิงวิธีการสำรวจโดยใช้การสุ่มตัวอย่าง แต่การที่เราจะได้ข้อมูลที่ดีนั้นย่อมหมายถึงว่าเราจะต้องสุ่มตัวแทนซึ่งเป็นตัวแทนอย่างแท้จริงมาจากประชากรด้วยเพราะถ้าเราได้กลุ่มตัวแทนที่ไม่ดี (ไม่เป็นตัวแทนที่แท้จริงของประชากร) จะทำให้ผลสรุปสู่ประชากรนั้นมีอคติได้

การสุ่มตัวอย่างแบ่งเป็น 2 แบบ

แบบที่ 1 การสุ่มตัวอย่างโดยไม่ใช้ทฤษฎีความน่าจะเป็น (Nonprobability Sampling)

แบบที่ 2 การสุ่มตัวอย่างโดยใช้ทฤษฎีของความน่าจะเป็น (Probability Sampling)

แบบที่ 1 การสุ่มตัวอย่างโดยไม่ใช้ทฤษฎีความน่าจะเป็น (Nonprobability Sampling) เป็นการสุ่มตัวอย่างที่ไม่ได้อิงหลักเกณฑ์ใด ๆ แต่จะใช้สิ่งที่ใกล้ตัวเป็นแนวทางในการเลือกตัวอย่างปัจจัยที่ใช้สามารถแยกได้ดังนี้คือ

1. การใช้ดุลยพินิจของผู้วิจัยเป็นหลัก วิธีนี้จำเป็นจะต้องอาศัยประสบการณ์ของผู้วิจัย ในเรื่องที่วิจัยอยู่ ตัวอย่างเช่น ผู้วิจัยเคยทำเรื่องทำนองนั้นมาบ่อย ๆ อยู่แล้ว ก็ย่อมจะมีประสบการณ์ในเรื่องของข้อมูลที่เกี่ยวข้องพอว่าลักษณะไหนเป็นตัวแทนของประชากร และอยู่ที่แหล่งไหน วิธีการนี้จะเหมาะสมกับกรณีที่ประชากรในเรื่องที่ทำการวิจัยนั้นมีกรอบตัวอย่าง (Sampling Frame)[1] ไม่สมบูรณ์พอ

2. การใช้ที่ตั้งของหน่วยประชากรเป็นเกณฑ์ตัวอย่างเช่น ถ้าเรานำ **แบบสอบถามจำนวนหนึ่ง** ไปสุ่มแจกให้กับคนที่กำลังชมภาพยนตร์อยู่โดยแจกเฉพาะคนที่อยู่แถวข้างประตู แถวข้างหน้าและแถวข้างโดยยึดถือความสะดวกของการเข้าออกเพื่อแจกแบบสอบถาม กรณีย์เช่นนี้เรียกว่าการเก็บข้อมูลโดยใช้ที่ตั้งของหน่วยเป็นเกณฑ์

3. การใช้ขนาดของหน่วยในประชากรเป็นเกณฑ์ ตัวอย่างเช่น สมมุติให้เลือกนักเรียนในชั้นหนึ่งมาเป็นตัวแทน 3 คน เราอาจจะเลือกคนที่โตที่สุดมา 1 คน คนตัวโตปานกลางมา 1 คน และคนที่เล็กที่สุดมา 1 คน ดังนี้เป็นต้น การเลือกโดยใช้เกณฑ์แบบ**นี้คือการใช้ขนาดเป็นเกณฑ์นั่นเอง**

4. การสุ่มแบบตามบุญตามกรรม วิธีนี้แล้วแต่ว่าผู้วิจัยจะสุ่มเลือกเองได้ เช่น การแทงข้าวสารในแต่ละกระสอบออกมาตรวจดูว่ามีข้าวเปลือกปนในปริมาณมากน้อยแค่ใด

5. การสุ่มโดยกำหนดโควต้า (Quota Sampling) วิธีนี้เป็นวิธีที่ใช้กันมากทางปฏิบัติ โดยเฉพาะอย่างยิ่งในการวิจัยทางด้านธุรกิจ วิธีนี้เป็นวิธีที่สะดวกทำได้รวดเร็วแต่ถึงอย่างไรก็มีปัญหาว่า เวลาจะวิเคราะห์ข้อมูล เราจะใช้ตัวสถิติอะไรเพราะวิธีนี้ไม่ใช้ Probability Sampling ซึ่งในสถานการณ์ การปฏิบัติจริงกลับปรากฏว่า ผู้เก็บข้อมูลเก็บโดยการใช้วิธี Quota Sampling แต่เวลาวิธีวิเคราะห์ข้อมูลกลับไปใช้การวิเคราะห์โดยใช้วิธีการของการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) มาใช้ซึ่งเป็นวิธีที่ขัดกัน

ตัวอย่างของการสุ่มตัวอย่างระบบแบบโควต้าคือ สมมุติว่าในการวิจัย เก็บข้อมูลจากนักศึกษาในมหาวิทยาลัยรามคำแหงผู้วิจัยตั้งหลักเกณฑ์ในการเก็บข้อมูลดังนี้คือ

สุ่มตัวอย่างจากนักศึกษาชั้นปีที่ 4	200	คน
สุ่มตัวอย่างจากนักศึกษาชั้นปีที่ 3	200	คน

---

[1] กรอบตัวอย่าง (Sampling Frame) หมายถึงรายชื่อของหน่วย (Unit) ต่าง ๆ ในประชากรพร้อมทั้งที่อยู่ของหน่วยตัวอย่าง เช่น ถ้าเราจะศึกษาถึงผู้ใช้โทรศัพท์ในเขตกรุงเทพฯ กรอกตัวอย่างในที่นี้คือ สมุดรายชื่อผู้ใช้โทรศัพท์ในกรุงเทพฯ

สุ่มตัวอย่างจากนักศึกษาชั้นปีที่ 2	300	คน
สุ่มตัวอย่างจากนักศึกษาชั้นปีที่ 1	400	คน
รวมนักศึกษาทั้งสิ้น	1100	คน

เมื่อตั้งเกณฑ์ในการกำหนดขนาดตัวอย่างได้แล้วก็ลงมือส่งแบบสอบถาม โดยการนำแบบสอบถามไปแจกให้กับนักศึกษาที่เดินเข้าประตูมหาวิทยาลัย สมมุติว่าเราทำอยู่จนกระทั่งได้ข้อมูลดังนี้คือได้ข้อมูลจากนักศึกษาชั้นปีที่ 4 ปีที่ 3 และปีที่ 1 ครบตามจำนวนต้องการแล้ว ถ้าพบนักศึกษารายถัดไปที่จะแจกแบบสอบถามอยู่ในชั้นปี 4 หรือปี 3 หรือปี 1 เราก็จะไม่แจกแบบสอบถามให้ แต่ถ้าเป็นนักศึกษาชั้นปีที่ 2 เราก็จะแจกให้ต่อไปจนครบ 300 คน ตามที่ตั้งโควต้าไว้แล้ว

แบบที่ 2 การสุ่มตัวอย่างโดยใช้ทฤษฎีของความน่าจะเป็น (Probability Sampling) เป็นการสุ่มตัวอย่างที่เปิดโอกาสให้แต่ละหน่วยในประชากรมีสิทธิ์ที่จะถูกรับเลือกให้เป็นหน่วยตัวอย่างซึ่งแตกต่างจากแบบที่ 1 ตรงที่บางหน่วยของประชากรไม่มีโอกาสได้รับเลือกเป็นหน่วยตัวอย่าง (Sampling Unit) ก็ได้การสุ่มตัวอย่างโดยวิธีนี้แบ่งได้เป็นหลายแบบแผนดังนี้คือ

2.1 การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) เป็นวิธีการสุ่มตัวอย่างโดยเปิดโอกาสให้แต่ละหน่วยในประชากรมีโอกาสได้รับเลือกเป็นตัวอย่างด้วยโอกาสเท่า ๆ กัน การสุ่มตัวอย่างแบบง่ายยังสามารถแบ่งได้เป็น 2 แบบย่อยคือ

2.1.1 การสุ่มตัวอย่างแบบคืนกลับ (Simple Randon Sampling with Replacement) กระบวนการสุ่มตัวอย่างนี้ก็คือการให้หน่วยที่ถูกรับเลือกแล้วกลับไปยังประชากรอีก ดังนั้นหน่วยตัวอย่างที่ถูกรับเลือกแล้วจึงมีโอกาสที่จะถูกรับเลือกในครั้งต่อ ๆ ไปอีกด้วย

2.1.2 การสุ่มตัวอย่างแบบไม่ใส่คืน (Simple Random Sampling without Replacement) การสุ่มตัวอย่างแบบนี้หมายความว่าเมื่อเราได้หน่วยใดหน่วยหนึ่งในประชากรแล้วเราจะไม่ใส่คืนกลับไปในประชากรอีก ดังนั้นหน่วยที่ถูกเลือกมาแล้วจึงไม่มีโอกาสได้รับเลือกอีก ซึ่งในสถานการณ์การปฏิบัติจริง ๆ มักเลือกใช้วิธีไม่ใส่คืน แต่ถ้าจะพิจารณากันให้ละเอียดแล้วในกรณีย์ที่ประชากรมีขนาดใหญ่มาก ๆ แล้ว การสุ่มโดยใส่คืนกับการสุ่มโดยไม่ใส่คืน จะเป็นวิธีการอันเดียวกันนั่นเอง

เครื่องมือที่ใช้นำมาช่วยในการสุ่มตัวอย่างโดยวิธีนี้มีอยู่ 2 แบบคือ

แบบที่ 1 ใช้วิธีจับฉลาก โดยการเขียนชื่อของแต่ละหน่วยในประชากร ซึ่งมีทั้งหมด N หน่วย แต่การจะเขียนชื่อของแต่ละหน่วยในประชากร แล้วนำมาจับฉลากจะประ-

สบปัญหายุ่งยาก ถ้าหากมีประชากรจำนวนมาก จะทำให้เสียเวลา จึงนิยมใช้เครื่องมือในแบบที่ 2

แบบที่ 2 โดยการใช้ตารางเลขสุ่มซึ่งตารางเลขสุ่มนี้จะมีอยู่ตอนท้ายของหนังสือสถิติแทบทุกเล่ม วิธีการใช้ก็สะดวกสบายเราสามารถกำหนดหน่วยตัวอย่างได้เลย

จะขอยกตัวอย่างการใช้ตารางเลขสุ่มดังนี้ สมมุติว่ามีประชากรอยู่ทั้งหมด 100 หน่วย เราต้องการสุ่มตัวอย่างมา 10 หน่วย โดยวิธีแบบสุ่มตัวอย่างง่าย

วิธีการ 1. กำหนดให้ประชากรคือ

$$Y_1, Y_2, \ldots\ldots\ldots\ldots, Y_{100}$$

หรือจะเขียนได้ในรูป $Y_{001}, Y_{002}, \ldots\ldots\ldots\ldots, Y_{100}$

โดยที่เลขที่เป็น subscript ของ Y คือหน่วยตัวอย่างที่เท่าใด เช่น Y_{001} คือหน่วยที่ 1 ในประชากร เป็นต้น

2. นำตารางเลขสุ่มมาสุ่มหน้า ยกตัวอย่างเช่นเราได้หน้าที่ 2 ก็ใช้หน้าที่ 2 แล้วสุ่มเลขสุ่มในหน้าที่ 2 สุ่มเลขมา 2 หลัก หลักแรกก็คือ แถว (row) และเลขหลักหลังคือสดมภ์ (Column) เช่นเราได้เลข 19 (ดูเลขในตารางเลขสุ่มหน้า 2 ประกอบ)

เลข 96 ที่ได้หลักแรกคือ 9 เราถือว่าเป็นค่าของแถว (row) และ หลักหลังคือ 6 เราถือเป็นสดมภ์ (column) เพื่อจะชี้เลขเริ่มต้นที่อ่านข้อมูลในที่นี้เราได้เลขเริ่มต้นคือ (อ่านจากแถวที่ 1 สดมภ์ที่ 9) คือเลข 7 ให้อ่านเลขมาทีละ 3 หลัก เราจะได้เลขดังนี้คือ 757, 687, 649, 020, 971, 877, 499, 042, 912, 272, อ่านเลขมา 10 จำนวน โดยที่แต่ละจำนวนมี 3 หลัก หลังจากนั้นจึงนำค่าขนาดประชากรเท่ากับ 100 ไปหาร เศษที่ได้จากการหารก็คือหน่วยที่จะเป็นตัวอย่างที่เราต้องการนั่นเอง ในตัวอย่างนี้จะได้ 057, 087, 049, 020, 071, 077, 099, 042, 012, 072

ในกรณีย์ที่เกิดหน่วยซ้ำกันแต่ผู้เก็บข้อมูลต้องการ การสุ่มตัวอย่างชนิดไม่ซ้ำ (Simple Random Sampling Without Replacement) เราก็ตัดหน่วยที่ซ้ำกันออก แล้วอ่านเลขสุ่มในลำดับถัดไป แล้วดำเนินการแบบเดียวกับที่แสดงมาให้ดูจนครบจำนวนหน่วยตัวอย่างที่ต้องการ

หมายเหตุ ใช้ตารางเลขสุ่มในหนังสือ

Introductory Statistics ; A Decision Map

โดย Thad R. Harshbarger

Table A-3 *Random digits**

00000	10097	32533	74520	13586	34573	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
00002	08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
00003	99019	02.529	09376	70715	38311	31165	88676	74397	04436	27659
00004	12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
00005	66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	35303	42614	86199	07439	23403	09732
00007	85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	60935	20344	25273	88435
00010	98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
00012	83452	99634	06288	98083	13746	70078	18475	40610	28711	77817
00013	88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
00015	65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
00017	74350	99817	77402	77214	43236	00210	43521	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
00020	91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
00024	63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
00025	61196	90446	26457	47774	51924	33729	65394	59593	42582	60527
00026	15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
00027	94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
00028	42481	16213	97344	08721	16868	48767	03071	12059	25701	46670
00029	23523	78317	73208	89837	68935	91416	26252	29663	05522	82562
00030	04493	52494	75246	33824	45862	51025	61962	79335	65337	12472
00031	00549	97654	64051	88159	96119	63896	54692	82391	23287	29529
00032	35963	15307	26898	09354	33351	35462	77974	50024	90103	39333
00033	59808	08391	45427	26842	83609	49700	13021	24892	78565	20106
00034	46058	85236	01390	92286	77281	44077	93910	83647	70617	42941
00035	32179	00597	87379	25241	05567	07007	86743	17157	85394	11838
00036	69234	61406	20117	45204	15956	60000	18743	92423	97118	96338
00037	19565	41430	01758	75379	40419	21585	66674	36806	84962	85207
00038	45155	14938	19476	07246	43667	94543	59047	90033	20826	69541
00039	94864	31994	36168	10851	34888	81553	01540	35456	05014	51176
00040	98086	24826	45340	28404	44999	08896	39094	73407	35441	31880
00041	33185	16232	41941	50949	89435	48581	88695	41994	37548	73043
00042	80951	00406	96382	70774	20151	23387	25016	25298	94624	61171
00043	79752	49140	71961	28296	69861	02591	74852	20539	00387	59579
00044	18633	32537	98145	06571	31010	24674	05455	61427	77938	91936

Source: Rand (1955), 1-4. Used by permission of the publisher.

Table A.-3 *Random digits (continued)*

00045	74029	43902	77557	32270	97790	17119	52527	58021	80814	51748
00046	54178	45611	80993	37143	05335	12969	56127	19255	36040	90324
00047	I1664	49883	52079	84827	59381	71539	09973	33440	88461	23356
00048	48324	77928	31249	64710	02295	36870	32307	57546	15020	09994
00049	69074	94138	87637	91976	35584	04401	10518	21615	01848	76938
000.50	09 I	88 20097	32825	39527	04220	86304	83389	a7374	64278	58044
0005I	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
0005 3	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813
00055	08358	69910	78542	42785	13661	58873	04618	97553	31223	08420
00056	28306	03264	81333	10591	40510	07893	32604	60475	94119	01840
00057	53840	86233	81594	13628	51215	90290	28466	68795	77762	20791
00058	91757	53741	61613	62269	50263	90212	55781	76514	83483	47855
00059	89415	92694	00397	58391	12607	17646	48949	72306	94541	37408
00060	77513	03820	86864	29991	68414	82774	51908	13980	72x93	55507
0006I	19502	37174	69979	20288	55210	29773	74287	75251	65344	67415
00062	21818	59313	93278	Xl757	05686	73156	07082	8046	31853	38452
00063	51474	66409	68107	23621	94049	91345	42836	09191	08007	45449
00064	99559	68331	6135	24176	69771	12830	74819	78142	43860	72834
00065	337 13	48007	93584	72869	51926	64721	58303	29822	93174	93972
00066	85374	86893	11303	22970	28834	34137'	73515	90400	71148	43643
00067	84133	89640	44035	52166	73852	70091	61222	60561	62327	18423
0006X	56732	16234	17395	96131	10123	91622	85496	57560	81604	18880
00069	65 138	56806	87648	85261	34313	65861	45875	21069	85644	47277
00070	38001	02176	81719	11711	71602	92937	74219	64049	65584	49698
00071	37402	96397	01304	77586	56271	10086	47324	62605	40030	37438
00072	97125	40348	87083	31417	21815	39250	75237	62047	15501	29578
00073	21826	41134	47143	34072	64638	85902	49139	06441	03856	54552
00074	73135	42742	95796	09035	85794	74296	08789	88156	64691	19202
00075	07638	77929	03061	18072	96207	44156	23821	99538	04713	66994
00076	60528	83441	07954	19814	59175	20695	05533	52139	61212	06455
00077	83596	35655	06958	92983	05128	09719	77433	53783	92301	50498
0007x	10x50	62746	99599	10507	13499	06319	53075	71839	06410	19362
00079	39x20	98952	43622	63147	64421	80814	43800	09351	31024	73167
00080	59.5 80	Oh4 78	75569	78800	88835	54486	23768	06156	04111	08408
0008 1	38508	07341	23793	48763	90822	97022	17719	04207	95954	49953
00082	30692	70668	94688	16127	56196	80091	82067	63400	05462	69200
00083	65443	95659	18288	27437	49632	24041	08337	65676	96299	90836
00084	27267	50264	13192	72294	07477	44606	17985	48911	97341	30358
00085	91307	0699 I	IY072	24210	36699	53728	28825	35793	28976	66252
00086	68434	94688	84473	13622	62126	98408	12843	82590	09815	93146
00087	48908	15877	54745	24591	35700	04754	83824	52692	54130	55160
00088	06913	45107	41672	78601	11883	09528	63011	98901	14974	40344
00089	10455	16019	14210	33712	91342	37821	88325	80851	43667	70883

Table A-3 *Random digits (continued)*

00090	12883	97343	65027	61184	04285	01392	17974	15077	90712	26769
00091	21778	30976	38807	36961	31649	42096	63281	02023	08816	47449
00092	19523	59515	65122	59659	86283	68258	69572	13798	16435	91529
00093	67245	52670	35583	16563	79246	86686	76463	34222	26655	90802
00094	60584	47377	07500	37992	45134	26529	26760	83637	41326	44344
00095	53853	41377	36066	94850	58838	73859	49364	73331	96240	43642
00096	24637	38736	74384	89342	52623	07992	12369	18601	03742	83873
00097	83080	12451	38992	22815	07759	51777	97377	27585	51972	37867
00098	16444	24334	36151	99073	27493	70939	85130	32552	54846	54759
00099	60790	18157	57178	65762	11161	78576	45819	52979	65130	04860
00100	03991	10461	93716	16894	66083	24653	84609	58232	88618	19161
00101	38555	95554	32886	59780	08355	60860	29735	47762	71299	23853
00102	17546	73704	92052	46215	55121	29281	59076	07936	27954	58909
00103	32643	52861	95819	06831	00911	98936	76355	93779	80863	00514
00104	69572	68777	39510	35905	14060	40619	29549	69616	33564	60780
00105	24122	66591	27699	06494	14845	46672	61958	77100	90899	15754
00106	61196	30231	92962	61773	41839	55382	17267	70943	78038	70267
00107	30532	21704	10274	12202	39685	23309	11061	68829	55986	66485
00108	03788	97599	75867	20717	74416	53166	35208	33374	87539	08823
00109	48228	63379	85783	47619	53152	67433	35663	52972	16818	60311
00110	60365	94653	35075	33949	42614	29297	01918	28316	98953	73231
00111	83799	42402	56623	34442	34994	41374	70071	14736	09958	18065
00112	32960	07405	36409	83232	99385	41600	11133	07586	15917	06253
00113	19322	53845	57620	52606	66497	68646	78138	66559	19640	99413
00114	11220	94747	07399	37408	48509	23929	27482	45476	85244	35159
00115	31751	57260	68980	05339	15470	48355	88651	22596	03152	19121
00116	88492	99382	14454	04504	20094	98977	74843	93413	22109	78508
00117	30934	47744	07481	84828	73788	06533	28597	20405	94205	20380
00118	22888	48893	27499	98748	60530	45128	74022	64617	82037	10268
00119	78212	16993	35902	91386	44372	15486	65741	14014	87481	37220
00120	41849	84547	46850	52326	34677	58300	74910	64345	19325	81549
00121	46352	33049	69248	93460	45305	07521	61318	31855	14413	70951
00122	11087	96294	14013	31792	59747	67277	76503	34513	39663	77544
00123	52701	08337	56303	87315	16520	69676	11654	99893	02181	68161
00124	57275	36898	81304	48585	68652	27376	92852	55866	88448	03584
00125	20857	73156	70284	24326	79375	95220	01159	63267	10622	48391
00126	15633	84924	90415	93614	33521	26665	55823	47641	86225	31704
00127	92694	48297	39904	02115	59589	49067	66821	41575	49767	04037
00128	77613	19019	88152	00080	20554	91409	96277	48257	50816	97616
00129	38688	32486	45134	63545	59404	72059	43947	51680	43852	59693
00130	25163	01889	70014	15021	41290	67312	71857	15957	68971	11403
00131	65251	07629	37239	33295	05870	01119	92784	26340	18477	65622
00132	36815	43625	18637	37509	82444	99005	04921	73701	14707	93997
00133	64397	11692	05327	82162	20247	81759	45197	25332	83745	22567
00134	04515	25624	95096	67946	48460	85558	15191	18782	16930	33361

Table A-3 *Random digits (continued)*

00135	83761	60873	43253	84145	60833	25983	01291	41349	20368	07126
00136	14387	06345	80854	09279	43529	06318	38384	74761	41196	37480
00137	51321	92246	80088	77074	88722	56736	66164	49431	66919	31678
00138	72472	00008	80890	18002	94813	31900	54155	83436	35352	54131
00139	05466	55306	93128	18464	74457	9056)	72848	11834	79982	68416
00140	39528	72484	82474	25593	48545	35247	18619	13674	18611	19241
00141	81616	18711	53342	44276	75122	11724	74627	73707	58319	15997
00142	07586	16120	82641	22820	92904	13141	32392	19763	61199	67940
00143	90767	04345	13574	17200	69902	63742	78464	22501	18627	90872
00144	40188	28193	29593	88627	94972	11598	62095	36787	00441	58997
00145	34414	82157	86887	55087	19152	00023	12302	80783	32624	68691
00146	63439	75363	44989	16822	36024	00867	76378	41605	65961	63488
00147	67049	09070	93399	45547	94458	74284	05041	49807	20288	34060
00148	79495	04146	52162	90286	54158	34243	46978	35482	59362	95938
00149	91704	30552	04737	21031	75051	93029	47665	64382	99782	93478
00150	94015	46874	32444	48277	59820	96163	64654	25843	41145	42820
00151	74108	88222	88570	74015	25704	91035	01755	14750	48968	38603
00152	62880	87873	95160	59221	22304	90314	72877	17334	39283	04149
00153	11748	12102	80580	41867	17710	59621	06554	07850	73950	79552
00154	17944	05600	60478	03343	25852	58905	57216	39618	49856	Y9326
00155	66067	42792	95043	52680	46780	56487	09971	59481	37006	22186
00156	54244	91030	45547	70818	59849	96169	61459	21647	87417	17198
00157	30945	57589	31732	57260	47670	07654	46376	25366	94746	49580
00158	69170	37403	86995	90307	94304	71803	26825	05511	12459	91314
00159	08345	88975	35841	85771	08105	59987	87112	21476	14713	71181
00160	27767	43584	85301	88977	29490	69714	73035	41207	74699	09310
00161	13025	14338	54066	15243	47724	66733	47431	43905	31048	56699
00162	80217	36292	98525	24335	24432	24896	43277	58874	11466	16082
00163	10875	62004	90391	61105	57411	06368\	53856	30743	08670	8474;
00164	54127	57326	26629	19087	24472	88779	30540	27886	61732	75454
00165	60311	42824	37301	42678	45990	43242	17374	52003	70707	70214
00166	49739	71484	92003	98086	76668	73209	59202	11973	02902	33250
00167	78626	51594	16453	94614	39014	97066	83012	09832	25571	77628
00168	66692	13986	99837	00582	81232	44987	09504	96412	90193	79568
00169	44071	28091	07362	97703	76447	42537	98524	97831	65704	09514
00170	41468	85149	49554	17994	14924	39650	95294	00556	70481	06905
00171	94559	37559	49678	53119	70312	05682	66986	34099	74474	20740
00172	41615	70360	64114	58660	90850	64618	80620	51790	11434	38072
00173	50273	93113	41794	86861	24781	89683	55411	85667	77535	99892
00174	41396	80504	90670	08289	40902	05069	95083	06783	28102	57816
00175	25807	24260	71529	78920	72682	07385	90726	57166	98884	08583
00176	06170	9796.5	88302	98041	21443	41808	68984	83620	89747	98882
00177	60808	54444	74412	81105	01176	28838	36421	16489	18059	51061
00178	80940	44893	10408	36222	80582	71944	92638	40333	67054	16067
00179	19516	90120	46759	71643	13177	55292	21036	82808	77501	97427

การใช้ตารางเลขสุ่ม เลือกตัวอย่างแบบง่ายนี้คือแม้ว่าจะสะดวกขึ้น แต่ถ้าหากประชากรมีมาก ๆ การสุ่มตัวอย่างโดยใช้วิธีนี้ก็ยังเสียเวลาอยู่มากดังนั้นเราอาจจะปรับปรุงวิธีการโดยใช้โปรแกรมคอมพิวเตอร์เลือกเลขสุ่มให้เราสะดวกขึ้น ดังตัวอย่างที่แสดงถึงการใช้ตารางเลขสุ่ม เลือกตัวอย่าง 10 หน่วย จากประชากร 100 หน่วย ถ้าจะใช้โปรแกรมดังต่อไปนี้ ก็จะสามารถเลือกตัวอย่างแบบง่ายได้เช่นกัน

ตัวอย่างโปรแกรมแสดงวิธีการเลือกตัวอย่าง (ใช้ภาษาเบสิค) และผลที่ได้

โปรแกรมที่ 1 เป็นการเลือกแบบเลือกแล้วใส่คืน ซึ่งอาจจะมีบางหน่วยที่ซ้ำกันเกิดขึ้น

```
 10 DIM X (10)
 20  FOR J = 1 TO 10
 30 LET Q = RND (-1)
 40 LET R = INT (Q*lOO) + 1
 50 LET X (J) = R
 60  NEXT J
 70 FORK = 1 TO 10
 80 PRINT X (K)
 90  NEXTK
100  END
```

ผลที่ได้

```
82
13
33
35
97
95
29
44
49
26
```

เลขที่พิมพ์ออกมาคือหน่วยที่เท่าใดที่ถูกคัดเลือกเป็นตัวอย่าง

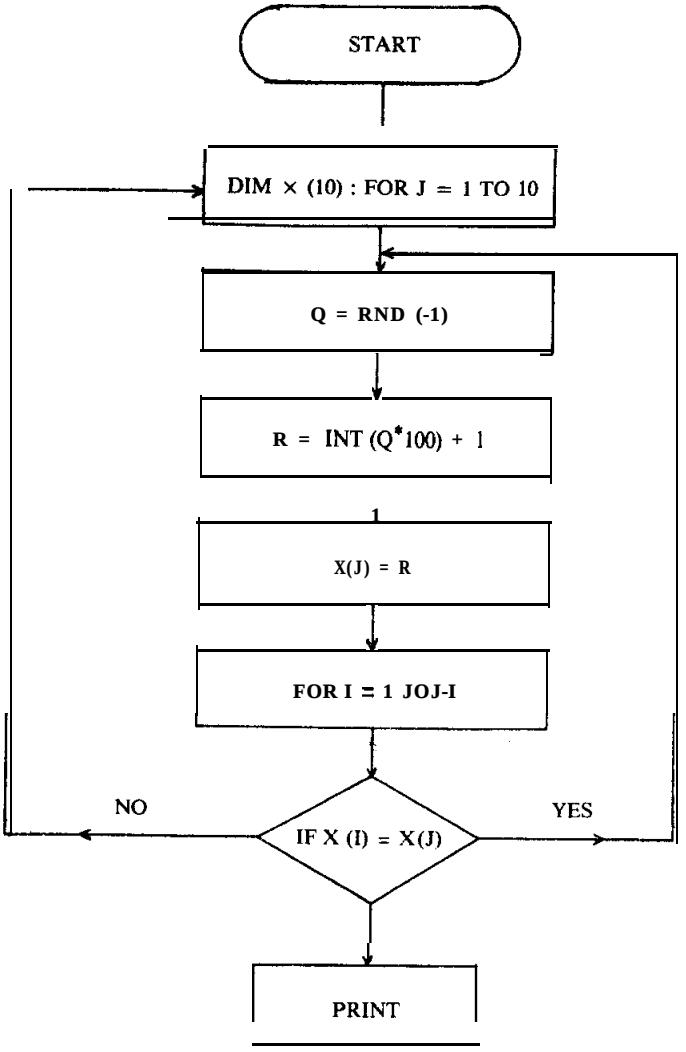
โปรแกรมที่ 2 เป็นการสุ่มตัวอย่างชนิดสุ่มแล้วไม่ใส่คืน กรณีย์นี้จะไม่มีหน่วยที่ซ้ำกันเกิดขึ้นเลย

```
10 DIM X (10)
20 FORJ = 1 TO 10
30 Q = RND (-1)
40 R = INT (Q*100) + 1
50 X (J) = R
55 IF J < > 1 THEN 70
60 IFJ = 1 THEN 95
70 FOR1 = 1 TO J-1
80 IF X(1) = X (J) THEN 30
90 NEXT I
95 NEXT J
100 FORK = 1 TO 10
110 PRINT X (K)
120 NEXT K
```

ผลที่ได้

```
59
34
35
75
 4
44
12
46
59
42
```

ผังโปรแกรมที่ 2



2. การสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) คือการสุ่มตัวอย่างชนิดที่เลือกหน่วยตัวอย่างเริ่มต้น (random start) มา 1 หน่วย แล้วเลือกต่อไปอีกทุก ๆ K หน่วย (fixed interval) การสุ่มตัวอย่างโดยวิธีนี้จะเป็นวิธีที่สะดวกตรงที่เมื่อเราเลือกได้หน่วยที่ 1 หน่วยอื่นอีก n-1 หน่วย (ขนาดตัวอย่างคือ n หน่วย) ก็จะมาโดยอัตโนมัติ ตัวอย่างเช่น สมมุติมีประชากรทั้งสิ้น 9 หน่วยในที่นี้คือ $Y_1, Y_2, \ldots\ldots, Y_9$ ถ้าเราเลือกตัวอย่างมา 3 หน่วย เราจะเลือกหน่วยตัวอย่างเริ่มต้นได้ดังนี้คือ ให้เลือกหน่วยตัวอย่างเริ่มต้น (random start) จากเลข 1 ถึงเลข $\frac{9}{3} = 3$

Rondom Start จะเป็นไปได้คือ 1,2,3

โดยที่ช่วงคงที่ (fixed interval) คือ 3 สมมุติเราเลือก random start ได้เป็น 2 ดังนั้นกลุ่มตัวอย่าง คือ y_2, y_{2+3}, y_{2+6}

กลุ่มตัวอย่างคือ $\{y_2, y_5, y_8\}$
แต่ถ้าได้ random start คือ 1
กลุ่มตัวอย่างคือ $\{y_1, y_4, y_7\}$
และถ้า random start คือ 3
กลุ่มตัวอย่างคือ $\{y_3, y_6, y_9\}$

จากตัวอย่างที่แสดงให้ดูจะเห็นว่า การสุ่มตัวอย่างแบบมีระบบทำให้ผู้เลือกเสียเวลาน้อย เพราะเพียงแต่เลือกหน่วยเริ่มต้นเท่านั้น หน่วยอื่นก็จะติดตามมาโดยอัตโนมัติ

กระบวนการเลือกตัวอย่างโดยวิธีนี้จะพิจารณาเห็นได้ว่า เป็นวิธีการที่เหมาะสมกับกรณีย์ที่หน่วยตัวอย่างตั้งอยู่ในสภาพที่เรียงลำดับกัน เช่น การสำรวจครัวเรือนที่ตั้งเรียงอยู่ในซอย โดยการใช้ที่ตั้งเป็นกรอบตัวอย่าง หรืออาจใช้ชื่อที่เรียงลำดับอยู่แล้ว เช่น สมุดโทรศัพท์เป็นกรอบตัวอย่าง หรือการสุ่มตัวอย่างหนังสือที่อยู่บนหิ้งเรียงกันอยู่ เป็นต้น จากตัวอย่างที่แสดงมาให้ดู ผู้อ่านก็คงพอที่จะนึกภาพออกว่า ในสถานะการณ์และสภาพเช่นใดที่สามารถจะใช้วิธีการสุ่มตัวอย่างแบบมีระบบ ข้อสังเกตที่จะพิจารณาก็คือค่าของ k ที่ใช้นั้นไม่จำเป็นจะต้องใช้เป็นหน่วยที่ห่างออก แต่อาจจะหมายถึงความยาวก็ได้ ตัวอย่างเช่น จากบัตรรายการหนังสือในชั้นที่มีอักษรตัว H ซึ่งบัตรรายการ เรียงอยู่มีความยาวของบัตรเรียงกันเท่ากับ 10 นิ้ว เราต้องการสุ่มหนังสือมาเพียง 5 เล่ม ตัว k ที่ใช้คือ $\frac{10}{5} = 2$ นิ้วก็ได้ ดังนั้น เราจะสุ่มตัวอย่างมา 1 ใบ จากใบที่ห่างไป 2 นิ้ว

การสุ่มตัวอย่างโดยวิธีนี้ ถึงแม้ว่าจะมีข้อดีตรงที่เสียเวลาน้อย และสามารถกระทำได้กับกรณีย์ที่ประชากรมีกรอบตัวอย่างไม่สมบูรณ์ โดยการใช้ที่ตั้งของหน่วยในประชากรเป็นแนวทางในการเลือก แต่ก็มีข้อเสียเช่นกัน ดังเช่น ถ้าหาหน่วยตัวอย่างในประชากรอยู่เรียงลำดับกันอย่างกระจัดกระจาย คำว่ากระจัดกระจายในที่นี้หมายถึงคุณลักษณะที่เราศึกษาแต่ละหน่วยนั้นไม่อยู่ใกล้เคียงกัน เช่น ถ้าเราศึกษาจำนวนสุกรที่เลี้ยงไว้ในครัวเรือนกลุ่มหนึ่งที่เราศึกษา โดยมีทั้งสิ้น 9 ครัวเรือน ตัวเลขที่แสดงไว้คือจำนวนสุกรที่เลี้ยงไว้ในแต่ละครัวเรือน

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
1	1	1	2	2	2	3	3	3

โดยที่ครัวเรือนที่ 1 มีสุกรเลี้ยงไว้ 1 ตัว

ครัวเรือนที่ 2 มีสุกรเลี้ยงไว้ 1 ตัว

⋮

ครัวเรือนที่ 9 มีสุกรเลี้ยงไว้ 3 ตัว

ถ้าเราจะสุ่มตัวอย่างครัวเรือนมาทั้งหมด 3 ครัวเรือน ดังนั้น $k = \frac{9}{3} = 3$

สมมุติเลือกได้ random start คือ 1 ดังนั้น กลุ่มตัวอย่างก็คือ y_1, y_6, y_9

และค่าที่จะนำไปใช้ในการวิเคราะห์ คือ {1, 2, 3}

ถ้าเราจะหาค่าเฉลี่ยจำนวนสุกรที่เลี้ยงในครัวเรือนโดยใช้กลุ่มตัวอย่างที่ได้ จะได้

$$x = \frac{1+2+3}{3} = 2 \text{ ตัว}$$ ซึ่งจะเห็นได้ว่าตรงกับค่าเฉลี่ยของประชากร

แต่ถ้าหากประชากรเรียงลำดับกันเป็นแบบที่กระจัดกระจายคือ 1,2,3,1,2,3, 1,2,3 ถ้าเราใช้วิธีการสุ่มตัวอย่างมา 3 หน่วย โดยที่ random start คือ 3 จะได้ค่าที่นำไปวิเคราะห์ คือ 3,3,3

ดังนั้นค่าเฉลี่ยจากกลุ่มตัวอย่างคือ $\frac{3+3+3}{3} = 3$

จะเห็นได้ว่าค่าที่ได้ผิดจากความจริงไป อันเนื่องมาจากเพราะการได้หน่วยตัวอย่างซึ่งมีลักษณะอย่างเดียวกันมาทั้งหมดนั่นเอง จะขอยกตัวอย่างเพื่อเทียบให้ดูดังนี้คือ บนถนนสายหนึ่งเราต้องการจะสุ่มร้านค้าต่าง ๆ มาเป็นตัวแทน สมมุติว่าบนถนนสายนี้ร้านค้าที่ขายสินค้าประเภทเดียวกันมิได้ตั้งเรียงติดกันเป็นกลุ่ม แต่ตั้งกระจัดกระจายในกรณีเช่นนี้เรา

อาจจะสุ่มได้ร้านค้าชนิดเดียวกัน เช่นร้านขายเครื่องสำอางค์มาเป็นตัวแทนทั้งหมดก็ได้ซึ่งไม่ใช่สิ่งที่เราต้องการ สิ่งที่เราต้องการคือได้ตัวแทนที่ขายสินค้าแต่ละประเภทต่าง ๆ กัน ดังนั้น ก่อนที่จะวางแผนการสุ่มตัวอย่างโดยวิธีนี้ก็ควรจะศึกษาคุณลักษณะการเรียงตัวของหน่วยต่าง ๆ ในประชากรเสียก่อน มิฉะนั้นอาจจะก่อให้เกิดผลเสียในการวิเคราะห์ข้อมูลก็ได้ ถ้าเกิดพบปัญหาอย่างที่กล่าวมาแล้ว

อนึ่งในการสุ่มตัวอย่างด้วยวิธีนี้ บางครั้งถ้าเกิดตัว k เป็นเลขไม่ลงตัวก็จะก่อให้เกิดปัญหาขึ้นได้เช่น ถ้าประชากรมีจำนวน 5 หน่วย เราเลือกมาเป็นตัวอย่างเพียง 2 หน่วย ดังนั้น $k = \frac{5}{2} = 2.5$ ค่าของ k ที่เป็นไปได้คือเลขจำนวนเต็ม ในที่นี้อาจจะเป็น 2 หรือ 3 ก็ได้

สมมติ k = 2 random start เลือกระหว่าง 1 กับ 2

ถ้า r = 1 กลุ่มตัวอย่างคือ $\{y_1, y_3, y_5\}$

r = 2 กลุ่มตัวอย่างคือ $\{y_2, y_4\}$

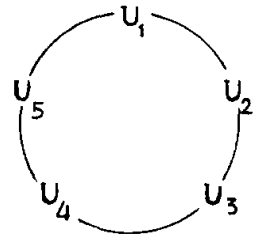
สมมุติ k = 3 random start เลือกระหว่าง 1,2 และ 3

ถ้า r = 1 กลุ่มตัวอย่างคือ $\{y_1, y_4\}$

r = 2 กลุ่มตัวอย่างคือ $\{y_1, y_5\}$

r = 3 กลุ่มตัวอย่างคือ $\{y_3\}$

กลุ่มตัวอย่างที่เกิดปัญหาคือ กลุ่มตัวอย่างเมื่อ r = 1, k = 2 และกลุ่มตัวอย่างที่ r = 3 และ k = 3 เพราะจะได้จำนวนตัวอย่างมากเกิน หรือน้อยกว่าที่ต้องการ ในที่นี้เราต้องการ n = 2 หน่วย ซึ่งผลกระทบจากกรณีย์นี้จะทำให้การประมาณค่าเกิดอคติขึ้นได้ ดังนั้นจึงมีผู้ปรับปรุงวิธีการเลือกตัวอย่างแบบชั้นภูมิเสียใหม่แทนที่จะเลือกเป็นเส้นตรง (Linear Systematic Samling) แบบที่ใช้อยู่ก็ปรับปรุงมาเป็นแบบวงกลม (Circular Sqstematic Sampling) ซึ่งมีวิธีการดังนี้ คือ นำประชากรมาเรียงกันเป็นวงกลมแทนที่จะเป็นเส้นตรง สมมุติว่า ประชากรประกอบด้วย U_1, U_2, U_3, U_4, U_5 นำมาเรียงกันเป็นวงกลมดังนี้คือ



กำหนดตัว N = 5

เลือกตัวอย่าง n = 2

$k = \frac{5}{2} = 2.5$ ค่าของ k ที่ได้เป็นไปได้คือ 2 หรือ 3

ถ้า k = 2 ให้เลือก random start ตั้งแต่ 1 ถึง N = 5 สมมุติเลือกได้

r = 1 กลุ่มตัวอย่างคือ $\{U_1, U_{1+2}\}$

r = 2 กลุ่มตัวอย่างคือ $\{U_2, U_{2+2}\}$

ถ้า r = 3 กลุ่มตัวอย่างคือ $\{U_3, U_{3+2}\}$

r = 4 กลุ่มตัวอย่างคือ $\{U_4, U_{4+2} \text{ mpd } 5\}$

r = 5 กลุ่มตัวอย่างคือ $\{U_5, U_{5+2} \text{ mod } 5\}$

วิธีการเลือกแบบวงกลมมีกฎเกณฑ์แตกต่างไปจากแบบเส้นตรงคือ

1. random start เลือกได้ตั้งแต่ 1 ถึง N (N คือขนาดประชากร)

2. เมื่อเลือกได้หน่วยที่ r เป็นหน่วยที่เริ่มต้นแล้วให้เลือกต่อไปอีกทุก ๆ k ดังนี้คือ $y_r, y_{r+1k}, y_{r+2k}, \ldots$ โดยที่เลือกจำนวนตัวอย่างเพียง n หน่วยเท่านั้นพอ ในกรณีย์ที่หน่วย r + ik เกินค่า N ให้ใช้หน่วยที่ r + ik mod N แทนตัวอย่างที่ r = 4 และ k = 2 หน่วยที่ 2 คือ 4 + 2 mod 5 = 6 mod 5 = 1 นั้น จะได้กลุ่มตัวอย่างดังกล่าวคือ

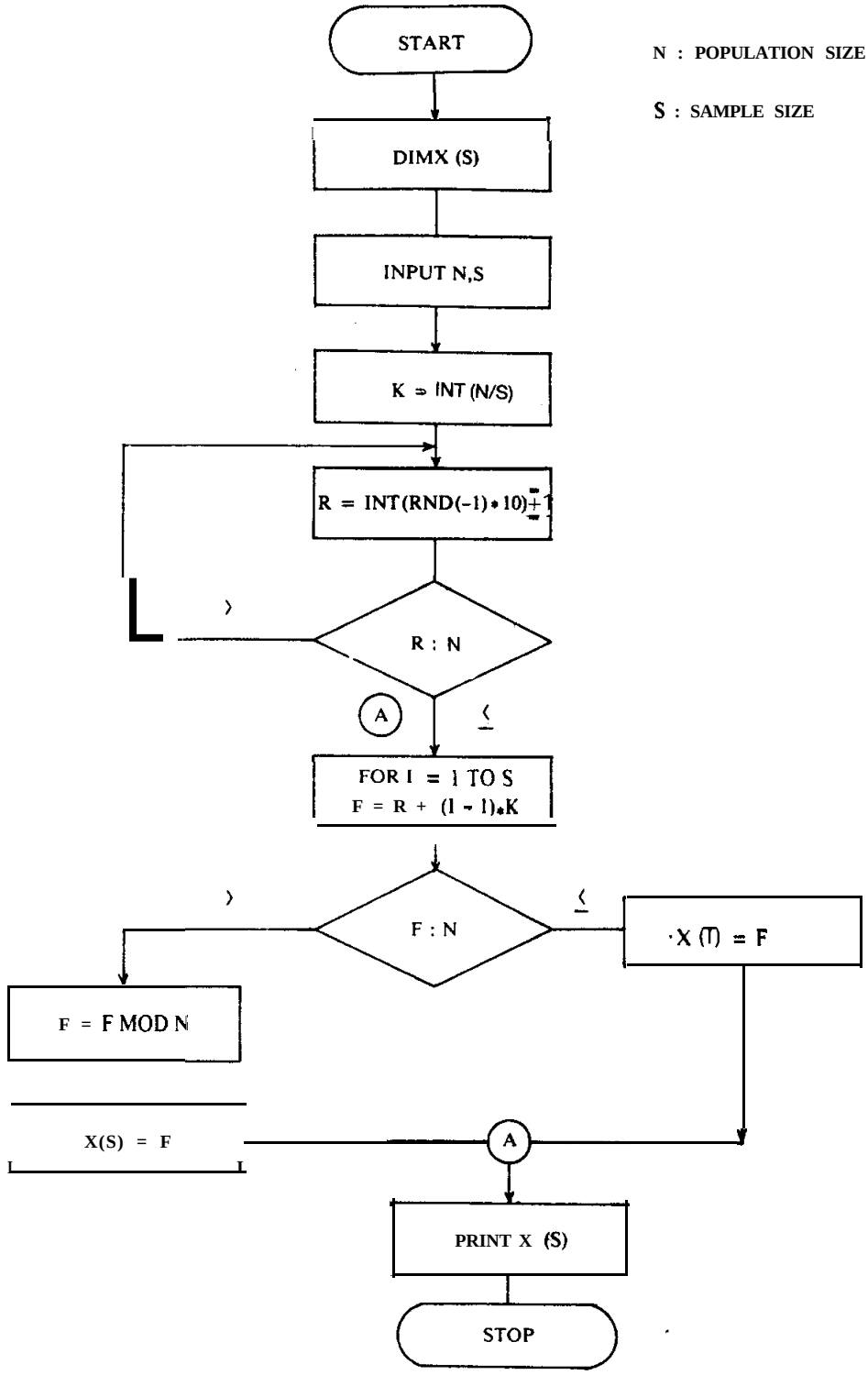
$\{U_4, U_1\}$

และเมื่อ r = 5 และ k = 3 หน่วยที่ 2 คือ หน่วยที่ 5 + 2 mod 5 = 7 mod 5 = 2 ดังนั้นจะได้กลุ่มตัวอย่างดังกล่าวคือ

$\{U_5, U_2\}$

โดยวิธีการใช้สุ่มแบบมีระบบโดยวิธีวงกลมจะช่วยแก้ไขของแบบวิธีเส้นตรงในข้อที่ได้จำนวนตัวอย่างมากไปหรือน้อยไปจากที่กำหนดไว้

ตัวอย่างโปรแกรมเลือกกลุ่มตัวอย่างโดยวิธีการสุ่มแบบมีระบบ (โดยวิธีวงกลม)



```
10    REM N : Population Size, S = Sample Size
20    DIM X (S)
30    INPUT N,S
40    K = INT (N/S)
50    R = INT (RND (-1) * N) + 1
60    IF R > K THEN 50
70    FORJ = 1 TO S
80    F = R + (J-l) * K
90    IF F > N THEN 120
100   X(J) = F
110   GO TO 70
120   F = F M O D N
130   X(J) = F
140   NEXT J
150   MAT PRINT X
```

การสุ่มตัวอย่างแบบแบ่งเป็นชั้นภูมิ (Stratified Sampling)

ความหมายของการสุ่มตัวอย่างแบบนี้ ก็คือการแบ่งประชากรออกเป็นชั้นภูมิ (Strata) โดยที่แต่ละชั้นภูมินั้นหน่วย (Unit) ในนั้นจะไม่มีการซ้ำ (lap) กันเลย เหตุผลที่เราแบ่งประชากรออกเป็นชั้นภูมิก็เพราะผู้สุ่มตัวอย่างไม่แน่ใจว่า ถ้าใช้แผนการสุ่มตัวอย่างแบบอื่น อาจจะได้ตัวแทนไม่ครอบคลุมประชากร เช่น ได้มาจากกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ เช่นสุ่มตัวอย่างจากประชากรในกรุงเทพฯ เพื่อศึกษาถึงเรื่องภาวะการดำรงชีพ เราอาจจะได้เฉพาะกลุ่มประชากรที่มีรายได้น้อยเท่านั้น ซึ่งไม่กระจายไปยังทุกกลุ่มของประชากรที่ต้องการศึกษา ดังนั้นผู้ศึกษาจึงแบ่งประชากรออกเป็นกลุ่มหรือที่เรียกว่าชั้นภูมิ (Strata) การแบ่งกลุ่มโดยปกติมักจะยึดถือเอาตัวแบ่งที่สำคัญในเรื่องที่ศึกษาเป็นกฎเกณฑ์ในการแบ่ง เช่น ถ้าเราศึกษาเรื่องผลผลิตของเกษตรกรในประเทศโดยเราจะแบ่งเกษตรกรในประเทศออกเป็น 4 ชั้นภูมิ โดยยึดเอาเขตที่ตั้งในแต่ละภาคเป็นหลัก จะได้ 4 ชั้นภูมิดังนี้ ชั้นภูมิที่ 1 คือเกษตรกรที่อยู่ในภาคเหนือ ชั้นภูมิที่ 2 คือเกษตรกรในภาคตะวันออกเฉียงเหนือ ชั้นภูมิที่ 3 คือเกษตรกรในภาคกลาง และชั้นภูมิที่ 4 คือเกษตรกรที่อยู่ในภาคใต้ หรือการศึกษาเรื่องความสัมพันธ์ระหว่างรายได้กับการศึกษา ผู้วิจัยอาจแบ่งประชากรที่ศึกษาออกเป็นชั้นภูมิโดยใช้รายได้เป็นตัวแปรในการแบ่งประชากร หรืออาจจะใช้ตัวแปรการศึกษาเป็นตัวแบ่งก็ได้ จากตัวอย่างที่แสดงให้เห็นนี้ผู้อ่านคงพอจะเข้าใจถึงวิธีการเลือกตัวแปรมาเป็นตัวกำหนดในการแบ่งชั้นภูมิได้ หลังจากที่แบ่งประชากรออกเป็นชั้นภูมิแล้วก็สุ่มตัวอย่างจากแต่ละชั้นภูมิมาจนครบเกณฑ์ที่ตั้งไว้ การสุ่มตัวอย่างจากแต่ละชั้นภูมินั้นก็จะประกอบด้วย 2 ขั้นตอนคือ

ขั้นตอนที่ 1 แบ่งกลุ่มตัวอย่างออกเป็นกลุ่มตัวอย่างย่อย จำนวนกลุ่มเท่ากับชั้นภูมิที่แบ่งเช่นแบ่งประชากรออกเป็น h ชั้นภูมิ ดังนั้นในแต่ละชั้นภูมิจะมีประชากรย่อยอยู่ N, จำนวน $N_1, N_2, \ldots N_h$ โดยที่ $N = N_1 + N_2 + \ldots + N_h$ (N คือจำนวนหน่วยในประชากร และ N_i ; $i = 1,2, \ldots\ldots h$ คือจำนวนประชากรในชั้นภูมิที่ i)

แบ่งกลุ่มตัวอย่าง n ออกมาเป็นกลุ่มตัวอย่างย่อย h กลุ่ม ดังนี้คือ $n_1, n_2, \ldots\ldots n_h$
โดยที่ $n = n_1, + n_2, \ldots\ldots + n_h$

ขั้นตอนที่ 2 สุ่มตัวอย่างย่อย n_i จาก N_i ของชั้นภูมิที่ i ดังนี้เรื่อยไป 1, 2,.....h 1, 2,.....h การสุ่มตัวอย่างในขั้นนี้อาจใช้วิธีเลือกแบบการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) ถ้าเป็นเช่นนี้เรียกว่า Stratified Simple Random Sampling แต่ถ้าสุ่มตัวอย่างย่อยแบบมีระบบ (Systematie) จะเรียกการสุ่มตัวอย่างชนิดนี้ว่า Stratified Systmatic Sampling

จากหลักการที่อธิบายมาผู้อ่านคงพอเข้าใจว่าการสุ่มตัวอย่างแบบชั้นภูมิก็คล้ายกับการสุ่มตัวอย่างแบบที่ 1 และแบบที่ 2 ดังที่กล่าวมาแล้วเพียงแต่การสุ่มตัวอย่างแบบมีชั้นภูมินี้กระทำกับประชากรย่อย ๆ h กลุ่มเท่านั้น ปัญหาที่ตามมาก็คือจะมีหลักเกณฑ์ในการจัดแบ่งกลุ่มตัวอย่างย่อย ๆ $n_1, n_2,\ldots,n_h$ ได้อย่างไร หลักการที่จะแบ่งกลุ่มตัวอย่าง n ออกมาเป็นกลุ่มตัวอย่างย่อย $n_1, n_2,\ldots,n_h$ มีหลายวิธีด้วยกัน การที่จะเลือกวิธีใดขึ้นอยู่กับความเหมาะสมและประสบการณ์ของผู้วิจัย

การจัดสรรกลุ่มตัวอย่าง (Allocation of Sample Size) **แบ่งได้ 4 วิธีคือ**

1. การจัดสรรแบบตามใจชอบ (Arbitarily Allocation) วิธีการนี้จะจัดสรรตัวอย่างย่อยให้กับแต่ละชั้นภูมิตามใจชอบ ไม่มีหลักเกณฑ์ใด ๆ วิธีนี้เหมาะสมกับผู้ที่มีประสบการณ์ในการวิจัยชนิดนั้นมามากพอจนรู้ถึงลักษณะของประชากรที่ศึกษา จึงจะสามารถจัดสรรตัวอย่างย่อยได้

2. การจัดสรรแบบแบ่งให้เท่า ๆ กัน (Equal Allocation) วิธีการนี้มีข้อดีตรงที่คิดง่าย สะดวก โดยที่ $n_i = \frac{n}{h}$,i = 1,2,.....,h ตัวอย่างเช่นประชากรที่ศึกษาแบ่งออกเป็น 5 ชั้นภูมิ เราต้องการสุ่มตัวอย่างขนาด 100 จากกลุ่มประชากรนี้ ดังนั้นจะต้องจัดสรรตัวอย่างขนาดย่อยให้กับแต่ละชั้นภูมิได้ดังนี้คือ $Ni = \frac{100}{5} = 20$, i = 1.....5

3. การจัดสรรแบบสัดส่วน (Proportional Allocation) วิธีการนี้ขจัดข้อเสียของข้อ 2 ตรงที่ ถ้าแบ่งกลุ่มตัวอย่างแบบเท่า ๆ กัน ก็จะทำให้เกิดปัญหาได้ 2 กรณีย์ คือ

กรณีย์ที่ 1 บางชั้นภูมิอาจจะมีจำนวนหน่วย (Unit) ในประชากรย่อยไม่เพียงพอที่จะสุ่มได้ คือ N_i น้อยกว่า n_i

กรณีย์ที่ 2 ในบางชั้นภูมิที่มีหน่วยในประชากรย่อยมาก แต่เราสุ่มมาเท่ากับชั้นภูมิที่มีหน่วยในประชากรน้อย ทำให้ข้อมูลที่จะมาศึกษา สำหรับประชากรในชั้นภูมิที่มีมากนั้นไม่เพียงพอ ในการจะอนุมานสู่ประชากรย่อย ในชั้นภูมินั้น ๆ ก็ได้

หลักการในการจัดสรรแบบสัดส่วนมีดังนี้คือ

สมมุติประชากรมีอยู่ N หน่วยแบ่งเป็น h ชั้นภูมิโดยที่แต่ละชั้นภูมิมี $N_1, N_2,\ldots. N_h$ หน่วย ต้องการสุ่มตัวอย่าง n หน่วย โดยแบ่งออกเป็น $n_1, n_2,\ldots.n_h$ หน่วยย่อย ตามวิธี

การนี้เราจะจัดแบ่งได้ดังนี้ คือ $n_i = N_i \frac{n}{N}$;i = 1,2,.....,h

ตัวอย่างเช่น ในประชากรกลุ่มหนึ่งที่ศึกษา แบ่งออกเป็น 4 ชั้นภูมิ โดยที่แต่ละชั้นภูมิมีประชากรย่อยดังนี้คือ

N_1 = 1000 หน่วย

N_2 = 2000 หน่วย

N_3 = 3000 หน่วย

N_4 = 4000 หน่วย

N = 1000 + 2000 + 3000 + 4000 = 10,000

กำหนด n = 100 หน่วย จะจัดสรรขนาดตัวอย่างได้ดังนี้คือ

$n_1 = 1000 \times \frac{100}{10,000} = 10$ หน่วย

$n_2 = 2000 \times \frac{100}{10,000} = 20$ หน่วย

$n_3 = 3000 \times \frac{100}{10,000} = 30$ หน่วย

$n_4 = 4000 \times \frac{100}{10,000} = 40$ หน่วย

4. การจัดสรรตัวอย่างแบบ อุตมะ (Optimum Allocation) การจัดสรรตัวอย่างโดยวิธีนี้จะมีกฎเกณฑ์และข้อจำกัดเพิ่มขึ้นในเรื่องของค่าใช้จ่ายดังนี้ คือ

กำหนดให้ฟังชันในการใช้จ่ายเป็นดังนี้คือ

$$c = C_0 + \sum_{i=1}^{h} c_i n_i$$

โดยที่ C คือค่าใช้จ่ายทั้งหมด (Total cost) ในโครงการวิจัย

c_0 คือค่าใช้จ่ายคงที่ (fixed cost) เช่น ค่าใช้จ่ายเกี่ยวกับค่าเช่าสถานที่ ค่าจ้างพนักงานถาวร ช่วยงานวิจัย ฯลฯ เป็นต้น

c_i คือค่าใช้จ่ายในการเก็บข้อมูลต่อ 1 หน่วยในชั้นภูมิที่ i

n_i คือจำนวนหน่วยตัวอย่างในชั้นภูมิที่ i

ดังนั้น $\sum_{i=1}^{h} c_i n_i$ คือค่าใช้จ่ายผันแปร (Variable cost) อันเนื่องมาจากการเก็บตัวอย่างทั้งหมดในทุกชั้นภูมิในการวิจัยที่เรามีข้อจำกัดเรื่องค่าใช้จ่าย (c) และเพื่อให้ได้ข้อมูลมีความผันแปร (Variation) ต่ำ เราสามารถกำหนดสูตรในการจัดสรรขนาดตัวอย่างได้ดังนี้ คือ

$$^{1}n_i = n (Ni\ Si / \sqrt{ci}) \sum_{i=1}^{h} Ni\ Si / \sqrt{ci}) ; i = 1,2,.....h$$

การจัดสรรขนาดตัวอย่างโดยวิธีนี้จะได้ว่าเป็นไปตามงบประมาณที่มีอยู่ และทำให้ค่าความแปรปรวน ($V(X_{st})$) ต่ำที่สุดด้วย

จากสูตรที่แสดงให้ดูจะพิจารณาได้ว่ามีค่าของ Si ร่วมอยู่ด้วย ซึ่งเป็นที่น่าสงสัยว่าในกรณีย์ที่เรายังได้ได้เก็บข้อมูลในเรื่องดังกล่าวมาวิเคราะห์แล้วเราจะหาค่า Si มาได้อย่างไร กระบวนการในการนำค่า Si มาใส่ในสูตรดังกล่าว ก็โดยการนำค่าประมาณของ Si ในเรื่องเดียวกัน หรือคล้ายกับที่มีผู้อื่นทำมาก่อน แล้วมาประมาณใส่ลงไป หรืออาจจะนำข้อมูลที่มาจากขั้นตอนในการทดลองแบบสอบถาม (pretest) มาประมาณใส่ลงไปก็ได้

4. แผนสำรวจแบบกลุ่ม (Cluster Sampling) การสุ่มตัวอย่างแบบนี้คือ การแบ่งประชากรออกเป็นกลุ่มคล้ายกับชั้นภูมิแต่ละจะเรียกว่า กลุ่ม (Cluster) ในแต่ละกลุ่มอาจจะมีจำนวนหน่วยแตกต่างกันไปเช่น แบ่งประชากรออกเป็น L กลุ่ม แต่ละกลุ่มมีจำนวน $N_1, N_2,.....,N_L$ หน่วย

จากนั้นเราจึงสุ่มมา k กลุ่ม จำนวนหน่วยทั้งหมดที่รวมกันใน k กลุ่มที่ได้เรียกว่ากลุ่มตัวอย่างขนาด n วิธีการที่สุ่มตัวอย่างแบบนี้เรียกว่าสุ่มมาเพียงขั้นตอนเดียว ตัวอย่างเช่นต้องการศึกษาเกษตรกรในจังหวัดอยุธยา ประชากรคือเกษตรกรทุกคนในเขตจังหวัดอยุธยา เราอาจแบ่งจังหวัดอยุธยาออกมาเป็น 4 ตำบล แล้วสุ่มตัวอย่างมา 10 ตำบล ตัวอย่างของเกษตรกรที่ได้คือเกษตรกรทุกคนที่อยู่ใน 10 ตำบลที่สุ่มได้

การสุ่มตัวอย่างโดยใช้แบบแผนนี้บางครั้งผู้วางแผนอาจจะเลือกเป็นหลายลำดับ ก็ได้ แทนที่จะเลือกเพียงลำดับเดียวดังที่กล่าวมาแล้ว เช่น จากตัวอย่างที่ยกมาตอนต้น เมื่อ

---

[1] การพิสูจน์สูตรที่ใช้ดังกล่าว ดูได้จาก มนตรี พิริยะกุล การสำรวจด้วยกลุ่มตัวอย่าง

เราสุ่มตัวอย่างตำบลอยุธยามา 10 ตำบล แล้วเราแบ่งตำบลทั้ง 10 ที่ได้ออกมาเป็นหมู่บ้าน สมมุติว่า ได้ทั้งหมด 60 หมู่บ้าน เราทำการเลือกมา 20 หมู่บ้าน จาก 60 หมู่บ้าน ดังนั้น เกษตรกรทุกคนใน 20 หมู่บ้านที่เลือกได้ก็คือ กลุ่มตัวอย่างที่เราต้องการนั่นเอง กระบวนการเลือกตัวอย่างแบบนี้เป็นการสุ่มตัวอย่าง 2 ขั้นตอน โดยที่ขั้นตอนที่ 1 ตำบล ที่เลือกได้ เรียกว่า Primary Sampling Unit และ ขั้นตอนที่ 2 หมู่บ้านที่เลือกได้ เรียกว่า Secondary Sampling Unit นักวิจัยบางรายอาจใช้การเลือกมากขั้นกว่านี้ก็ได้ ขึ้นอยู่กับความเหมาะสมของหน่วยตัวอย่างที่ตั้งไว้