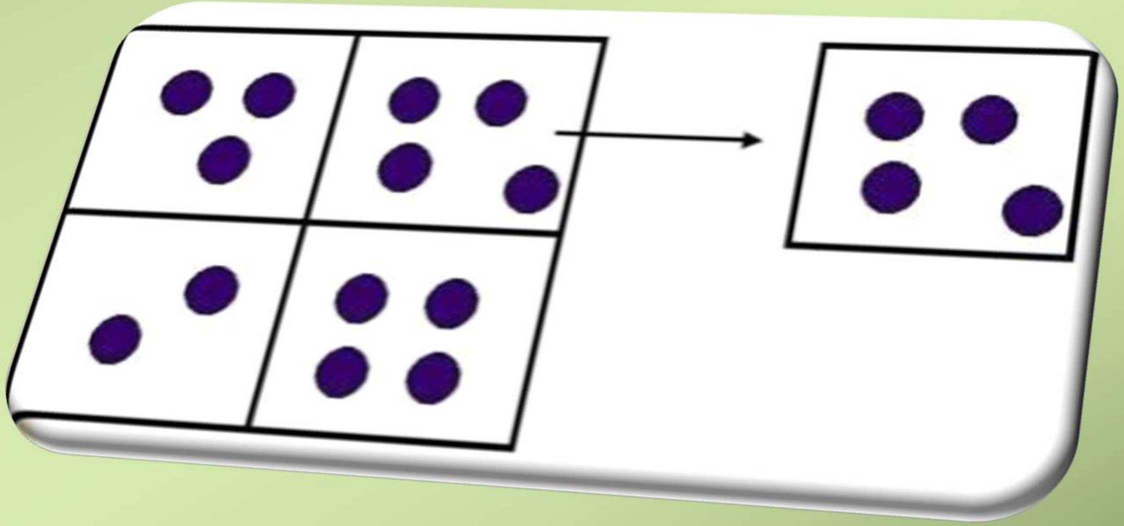
การสุ่มตัวอย่าง



ประเภทของการสุ่มตัวอย่าง

1. การสุ่มตัวอย่างโดยใช้ทฤษฎีความน่าจะเป็น (Probability Sampling) มีวิธีที่นิยมใช้ 4 วิธีดังนี้

1. การสุ่มตัวอย่างสุ่มอย่างง่าย (Simple random Sampling)

เป็นการสุ่มตัวอย่างจากประชากรทั้งหมดโดยให้แต่ละหน่วยมีโอกาสถูกเลือกเท่าๆกันทุกหน่วย สำหรับวิธีการเลือกตัวอย่างสามารถทำได้หลายวิธี แต่วิธีที่นิยมใช้กัน เช่น ใช้วิธีการจับฉลาก ซึ่งเป็นวิธีที่เหมาะกับขนาดตัวอย่างไม่มากนักและการใช้ตารางเลขสุ่ม

2. การสุ่มตัวอย่างแบบมีระบบ

(Systematic Sampling)

การสุ่มตัวอย่างแบบมีระบบ คือ การสุ่มตัวอย่างที่ทำการเลือกหน่วยตัวอย่างแรกแบบสุ่มจากหน่วยที่ 1 ถึง หน่วยที่ k โดยที่ k = และต่อจากนั้นจะเลือกหน่วยตัวอย่างต่อไปทุกๆ k หน่วย จนกระทั่งครบ n หน่วยตามที่ต้องการ กล่าวคือ ถ้าเลือกได้หน่วยตัวอย่างแรกเป็นหน่วยที่ i เมื่อ $1 \leq i \leq k$ หน่วยตัวอย่างที่จะถูกเลือกเป็นตัวอย่างคือ หน่วยตัวอย่างที่ i , i +k , i +2k ,i+3k , ... , i+(n – 1)k

3. การสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified random sampling)

การสุ่มตัวอย่างแบบแบ่งชั้นภูมิ เป็นการเลือกตัวอย่างจากประชากรที่มีการแบ่งออกเป็นชั้นภูมิ (stratum) ตามลักษณะบางอย่าง แล้วเลือกตัวแทนของประชากรในแต่ละชั้นภูมิขึ้นมาจำนวนหนึ่ง เพื่อเป็นตัวอย่างในการสำรวจ วิธีการแบ่งประชากรออกเป็นชั้นภูมิ เรียกว่า stratification แต่ละชั้นภูมิของประชากรที่แบ่งออกไปเรียกว่า stratum หลักสำคัญในการแบ่งก็คือ ให้หน่วยที่อยู่ในชั้นภูมิเดียวกันควรมีความคล้ายคลึงกัน (homogeneity within stratum) มากที่สุด แต่มีความแตกต่างกันระหว่างชั้นภูมิมากที่สุด (heterogeneity between stratum)

4. การสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster randoms sampling)

เป็นการสุ่มตัวอย่างโดยที่ประชากรอยู่รวมกันเป็นกลุ่ม ๆ (Cluster) โดยแต่ละกลุ่มมีลักษณะภายในกลุ่มที่หลากหลายหรือมีความแตกต่าง แต่ระหว่างกลุ่มมีความคล้ายคลึงกัน เช่น กลุ่มเกษตรในหมู่บ้าน กลุ่มนักเรียนในห้องเรียน เป็นต้น จำนวนของกลุ่มต่าง ๆ จะถูกสุ่มขึ้นมาทำการศึกษา เมื่อสุ่มได้กลุ่มใดก็จะนำสมาชิกที่อยู่ในกลุ่มนั้น ๆ ทั้งหมดมาทำการศึกษา เช่น การศึกษาเกี่ยวกับครัวเรือนในประเทศไทย เราอาจแบ่งครัวเรือนออกเป็นกลุ่มโดยใช้ตำบลเป็นหลัก แล้วทำการสุ่มตำบลเมื่อสุ่มตำบลแล้ว ก็ทำการเก็บรวบรวมข้อมูลจากทุกครัวเรือนที่อยู่ในตำบลที่สุ่มได้นั้น ๆ

2. การสุ่มตัวอย่างโดยไม่ใช้ทฤษฎีความน่าจะเป็น (Non Probability Sampling)

การเลือกตัวอย่างโดยไม่ใช้ทฤษฎีความน่าจะเป็น (Non Probability Sampling) เป็นการเลือกตัวอย่างที่เหมาะกับกรณีที่ไม่มีกรอบตัวอย่างหรือในกรณีที่กรอบตัวอย่างไม่สมบูรณ์ ข้อมูลมีการกระจายมาก ต้องการประหยัดเวลาและค่าใช้จ่ายในการเก็บรวบรวมข้อมูล โดยการเลือกตัวอย่างโดยไม่ใช้ทฤษฎีความน่าจะเป็นที่นิยมใช้จะมีอยู่ 3 วิธีคือ

1. การสุ่มตัวอย่างตามความสะดวก (Convenience Sampling)

เป็นการสุ่มตัวอย่างตามความสะดวกของผู้จัดเก็บข้อมูล เช่น โทรศัพท์ถามความเห็น การออกจดหมายส่งแบบสอบถาม เป็นต้น โดยการสุ่มตัวอย่างชนิดนี้จะไม่สามารถประมาณค่าพารามิเตอร์ของประชากรที่สนใจได้

2. การสุ่มตัวอย่างโดยใช้วิจารณญาณ (Judgment Sampling)

เป็นการเลือกตัวอย่างโดยตัดสินใจว่าจะเลือกตัวอย่างใดเป็นตัวแทนของประชากร โดยเจาะจงหน่วยตัวอย่างที่จะเลือกไว้แล้ว เช่น กำหนดเกณฑ์การศึกษากลุ่มโดยคัดเลือกดูจากขนาดครอบครัว รายได้ ฯลฯ สำหรับการเลือกตัวอย่างแบบนี้ผู้วิจัยจะต้องมีความรู้และประสบการณ์ รวมทั้งมีการวางแผนเป็นอย่างดีในการเลือกตัวอย่างขึ้นมาเป็นตัวแทนประชากร ถึงแม้ว่าการสุ่มตัวอย่างแบบนี้จะไม่สามารถบอกถึงระดับความผิดพลาดได้อย่างแน่ชัด แต่จะให้ผลดีกว่าการสุ่มตัวอย่างตามความสะดวก

3. การสุ่มตัวอย่างโดยการกำหนดโควตา (Quota Sampling)

การสุ่มตัวอย่างโดยการกำหนดโควตา เรียกอีกอย่างหนึ่งว่าการสุ่มตัวอย่างแบบกำหนดจำนวน ในการสุ่มแบบนี้ ประชากรจะถูกแบ่งออกเป็นกลุ่มตามลักษณะที่เลือกเอาไว้เป็น เพศ อายุ การศึกษา ฯลฯ โดยกำหนดสัดส่วนของแต่ละกลุ่ม เช่น จะศึกษาประชากร 300 คน ก็ให้ประมาณว่าจะใช้เพศชาย เพศหญิงอย่างละกี่คน ระดับการศึกษา อายุ รายได้ จำนวนกลุ่มละเท่าไร การจัดสัดส่วนระหว่างกลุ่มพยายามให้มีเท่ากันในแต่ละกลุ่ม แล้วจึงลงมือเก็บข้อมูลแบบใช้ความสะดวก convenience Sampling) คือเก็บเฉพาะคนที่ให้ความร่วมมือจนครบจำนวนตามต้องการ โดยข้อจำกัดของการสุ่มแบบนี้ คือ ผู้วิจัยจะไม่ทราบสัดส่วนที่แน่นอนของผู้ตอบแต่ละกลุ่ม และการคัดเลือกผู้ที่มีคุณสมบัติตรงตามกำหนด ในทางปฏิบัติเป็นไปได้ยาก

การกำหนดขนาดตัวอย่าง

ในการเก็บรวบรวมข้อมูลโดยการสำรวจด้วยตัวอย่าง ปัญหาที่พบเสมอก็คือ ควรจะใช้ขนาดตัวอย่างเท่าไร โดยทั่วไปหากใช้ตัวอย่างที่มีขนาดใหญ่จะทำให้การประมาณค่าหรือการทดสอบสมมติฐานมีความแม่นยำ แต่ในทางปฏิบัติจะมีข้อจำกัดในการกำหนดขนาดตัวอย่าง เช่น ระยะเวลา งบประมาณ เป็นต้น

สำหรับวิธีกำหนดขนาดตัวอย่าง แบ่งออกเป็น 2 ประเภท คือ ประเภทที่ 1 การใช้ตารางกำหนดขนาดตัวอย่าง และ ประเภทที่ 2 การคำนวณโดยใช้สูตร

ประเภทที่ 1 การใช้ตารางกำหนดขนาดตัวอย่าง

- ตารางของทาโร ยามาเน (Taro Yamane)
- ตารางของเครจซี่และมอร์แกน (R. V. Krejcie and D.W. Morgan)

ตารางกำหนดขนาดตัวอย่างของ Taro Yamane ที่ระดับความเชื่อมั่น 95%

ขนาดประชากร	ขนาดตัวอย่างตามความคลาดเคลื่อน					
	±1%	±2%	±3%	±4%	±5%	±10%
500	-	-	-	-	222	83
1,000	-	-	-	385	286	91
1,500	-	-	638	441	316	94
2,000	-	-	714	476	333	95
2,500	-	1,250	769	500	345	96
3,000	-	1,364	811	517	353	97
3,500	-	1,458	843	530	359	97
4,000	-	1,538	870	541	364	98
4,500	-	1,607	891	549	367	98
5,000	-	1,667	909	556	370	98
6,000	-	1,765	938	566	375	98
7,000	-	1,842	959	574	378	99
8,000	-	1,905	976	580	381	99
9,000	-	1,957	989	584	383	99
10,000	5,000	2,000	1,000	588	385	99
15,000	6,000	2,143	1,034	600	390	99
20,000	6,667	2,222	1,053	606	392	100
25,000	7,143	2,273	1,064	610	394	100
50,000	8,333	2,381	1,087	617	397	100
100,000	9,091	2,439	1,099	621	398	100
>100,000	10,000	2,500	1,111	625	400	100

ตารางกำหนดขนาดตัวอย่างของ Taro Yamane ที่ระดับความเชื่อมั่น 99%

ขนาดประชากร	ขนาดตัวอย่างตามความคลาดเคลื่อน				
	±1%	±2%	±3%	±4%	±5%
500	-	-	-	-	-
1,000	-	-	-	-	474
1,500	-	-	-	726	563
2,000	-	-	-	826	621
2,500	-	-	-	900	662
3,000	-	-	1364	958	692
3,500	-	-	1458	1003	716
4,000	-	-	1539	1041	735
4,500	-	-	1607	1071	750
5,000	-	-	1667	1098	763
6,000	-	2903	1765	1139	783
7,000	-	3119	1842	1171	798
8,000	-	3303	1905	1196	809
9,000	-	3462	1957	1216	818
10,000	-	3600	2000	1233	826
15,000	-	4091	2143	1286	849
20,000	-	4390	2222	1314	861
25,000	11842	4592	2273	1331	869
50,000	15517	5056	2381	1368	883
100,000	18367	5325	2439	1387	892
>100,000	22500	5625	2500	1406	900

ตัวอย่าง

ถ้าจำนวนประชากร 2,500 คน ที่ระดับความเชื่อมั่น 95% ความคลาดเคลื่อนที่ผู้วิจัยยอมรับได้ 3% จะใช้จำนวนตัวอย่างเพื่อการวิจัยในครั้งนี้เท่ากับ 769 คน ในกรณีที่จำนวนของประชากรไม่มีในตารางอาจจะใช้การประมาณขนาดตัวอย่างโดยเลือกใช้ขนาด ประชากรที่มีจำนวนใกล้เคียงที่สุดหรือใช้วิธีเปรียบเทียบสัดส่วน

ตัวอย่างตารางกำหนดขนาดตัวอย่างของเครจซี่และมอร์แกน

ขนาดประชากร	ขนาดตัวอย่าง	ขนาดประชากร	ขนาดตัวอย่าง	ขนาดประชากร	ขนาดตัวอย่าง
10	10	220	140	1,200	291
15	14	230	144	1,300	297
20	19	240	148	1,400	302
25	24	250	152	1,500	306
30	28	260	155	1,600	310
35	32	270	159	1,700	313
40	36	280	162	1,800	317
45	40	290	165	1,900	320
50	44	300	169	2,000	322
55	48	320	175	2,200	327
60	52	340	181	2,400	331
65	56	360	186	2,600	335
70	59	380	191	2,800	338
75	63	400	196	3,000	341
80	66	420	201	3,500	346
85	70	440	205	4,000	351
90	73	460	210	4,500	354
95	76	480	214	5,000	357
100	80	500	217	6,000	361
110	86	550	226	7,000	364
120	92	600	234	8,000	367
130	97	650	242	9,000	368
140	103	700	248	10,000	370
150	108	750	254	15,000	375
160	113	800	260	20,000	377
170	118	850	265	30,000	379
180	123	900	269	40,000	380
190	127	950	274	50,000	381
200	132	1,000	278	75,000	382
210	136	1,100	285	100,000	384

ประเภทที่ 2 การคำนวณจำนวนของตัวอย่างโดยใช้สูตร

ซึ่งมีหลายสูตร แต่ละสูตรมีเงื่อนไขแตกต่างกัน เช่น พารามิเตอร์ที่ผู้วิจัยต้องการวิเคราะห์ ขนาดประชากร ความคลาดเคลื่อนระหว่างค่าพารามิเตอร์กับ ค่าสถิติ ระดับความเชื่อมั่น ระดับนัยสำคัญ เป็นต้น โดยในบทนี้จะขอกล่าวเพียงวิธีที่นิยมใช้ 4 วิธี ดังนี้

1. เมื่อทราบจำนวนของประชากรและผู้วิจัยกำหนดความคลาดเคลื่อนระหว่างค่าพารามิเตอร์กับค่าสถิติคำนวณโดยใช้สูตรของทาโร ยามาเน ที่ระดับความเชื่อมั่น 95%

$$n = \frac{N}{1 + Ne^2}$$

เมื่อ n เป็น จำนวนของตัวอย่าง

N เป็น จำนวนของประชากร

e เป็น สัดส่วนของความคลาดเคลื่อนเทียบกับค่าพารามิเตอร์

ตัวอย่าง

ถ้าในงานวิจัยมีจำนวนประชากร 4,000 คน ความคลาดเคลื่อนที่ผู้วิจัยยอมรับได้ 5% จะใช้จำนวนตัวอย่างเพื่อการวิจัยครั้งนี้เท่าใด

วิธีทำ จากสูตร $n = \frac{N}{1 + Ne^2}$

$$= \frac{4,000}{1 + 4000(0.05^2)}$$

$= 363.63$ » 364 คน

2. เมื่อทราบจำนวนของประชากรและสัดส่วนตาม
ลักษณะที่ผู้วิจัยต้องการศึกษาจากกลุ่มประชากร
กำหนดความคลาดเคลื่อนระหว่างค่าพารามิเตอร์
กับค่าสถิติและระดับความเชื่อมั่น คำนวณโดยใช้
สูตรของเครจซี่และมอร์แกน

$$n = \frac{\chi^2 Np(1-p)}{e^2(N-1) + \chi^2 p(1-p)}$$

เมื่อ n เป็นจำนวนของตัวอย่าง

N เป็นจำนวนของประชากร

e เป็นสัดส่วนของความคลาดเคลื่อนเทียบกับค่าพารามิเตอร์

p เป็นสัดส่วนตามลักษณะที่ผู้วิจัยต้องการศึกษาจากกลุ่มประชากร

χ^2 เป็นค่าไคสแควร์ (Chi – Square) ที่องศาความเป็นอิสระเท่ากับ 1 และระดับความเชื่อมั่น $(1-\alpha)100\%$

ตัวอย่าง

ในการศึกษาภาวการณ์เป็นผู้นำของอาจารย์ในมหาวิทยาลัยแห่งหนึ่งที่มีอาจารย์จำนวน 600 คน โดยยอมให้เกิดความคลาดเคลื่อนได้ 5% และที่ระดับความเชื่อมั่น 95% จะต้องใช้ขนาดตัวอย่างเท่าใด ถ้าร้อยละ 50 ของคนทำงานในมหาวิทยาลัยแห่งนี้เป็นอาจารย์

วิธีทำ จากสูตร

$$n = \frac{\chi^2 Np(1-p)}{e^2(N-1) + \chi^2 p(1-p)}$$

จะได้ N = 600 , e = 0.05 , p = 0.5 และ $\chi^2_{1,0.95} = 3.841$

จะได้

$$n = \frac{3.841 \times 600 \times 0.5 \times 0.5}{0.05^2(600-1) + 3.841 \times 0.5 \times 0.5} = 234.42$$

ดังนั้น เพื่อให้เกิดความคลาดเคลื่อน 5% และระดับความเชื่อมั่น 95% จะต้องใช้ขนาดตัวอย่าง 234 ตัวอย่าง

3. ต้องการวิเคราะห์ค่าเฉลี่ยของประชากร โดยกำหนดความคลาดเคลื่อนระหว่างค่าเฉลี่ยของประชากรกับค่าเฉลี่ยของกลุ่มตัวอย่างไม่เกิน e และระดับนัยสำคัญ α หรือที่ระดับความเชื่อมั่น $(1 - \alpha)100\%$ สำหรับการวิเคราะห์สองด้าน

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{e} \right)^2$$

เมื่อ n เป็นจำนวนของตัวอย่าง

$Z_{\frac{\alpha}{2}}$ เป็นค่าปกติมาตรฐานที่ $P\left(Z^{2}Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$

σ เป็นส่วนเบี่ยงเบนมาตรฐานของประชากร

e เป็นค่าความคลาดเคลื่อนระหว่างค่าเฉลี่ยของประชากรกับค่าเฉลี่ยของกลุ่มตัวอย่างถ้าไม่ทราบค่าส่วนเบี่ยงเบน มาตรฐานของประชากรให้ประมาณด้วยส่วนเบี่ยงเบน มาตรฐานของกลุ่มตัวอย่าง

ตัวอย่าง

นักจิตวิทยาผู้หนึ่งตรวจพบว่า ส่วนเบี่ยงเบนมาตรฐานของระยะเวลาที่เกิดปฏิกิริยาตอบสนองเป็น 0.07 วินาที จะต้องใช้จำนวนตัวอย่างเท่าใดเพื่อทำให้มีความเชื่อมั่น 99% และมีค่าความคลาดเคลื่อนจากค่าเฉลี่ยไม่เกิน 0.02 วินาที

วิธีทำ จากโจทย์ทราบว่า $\sigma = 0.07$ $e = 0.02$ และ $Z_{\frac{\alpha}{2}} = Z_{0.05} = 2.575$

จากสูตร

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{e} \right)^2$$

$$= \left(\frac{(2.575)(0.07)}{0.02} \right)^2 = 81.2251$$

ดังนั้น เพื่อให้เกิดความเชื่อมั่น 99% ค่าความคลาดเคลื่อนไม่เกิน 0.02 วินาที จะต้องใช้ขนาดตัวอย่าง 81 ตัวอย่าง

4. ต้องการวิเคราะห์สัดส่วนของประชากร โดยกำหนดความคลาดเคลื่อนระหว่างสัดส่วนของประชากรกับสัดส่วนของกลุ่มตัวอย่างไม่เกิน e และระดับนัยสำคัญ α สำหรับการวิเคราะห์สองด้าน

$$n = \frac{Z^2_{\frac{\alpha}{2}}\hat{p}\hat{q}}{e^2}$$

เมื่อ n เป็นจำนวนของตัวอย่าง

$Z_{\frac{\alpha}{2}}$ เป็นค่าปกติมาตรฐานที่ $\frac{\alpha}{2}$

$\hat{p}$ เป็นสัดส่วนลักษณะที่ศึกษาของกลุ่มตัวอย่าง

$\hat{q}$ เป็นสัดส่วนลักษณะที่ไม่ได้ศึกษาของกลุ่มตัวอย่าง

ซึ่ง $\hat{q} = 1 - \hat{p}$

e เป็นค่าความคลาดเคลื่อนระหว่างสัดส่วนของประชากรกับสัดส่วนของกลุ่มตัวอย่าง

ตัวอย่าง

ต้องการประมาณสัดส่วนของประชาชนที่มีความพึงพอใจต่อการให้บริการของศูนย์วิทยาศาสตร์แห่งหนึ่ง ให้ความคลาดเคลื่อนไม่เกิน 15% ที่ระดับความเชื่อมั่น 95% ควรสุ่มตัวอย่างจำนวนเท่าใด จากข้อมูลเก่าทราบว่ามีร้อยละ 45 ของประชาชนที่มีความพึงพอใจต่อการให้บริการของศูนย์วิทยาศาสตร์

วิธีทำ จากโจทย์จะได้ $e = 0.15$ $\hat{p} = 0.45, \hat{q} = 0.55$ และ $Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$

จากสูตร

$$n = \frac{Z^2_{\frac{\alpha}{2}}\hat{p}\hat{q}}{e^2} = \frac{(1.96)^2(0.45)(0.55)}{(0.15)^2}$$

ดังนั้น การประมาณสัดส่วนของประชาชนที่มีความพึงพอใจต่อการให้บริการของศูนย์วิทยาศาสตร์ควรสุ่มสอบถามจากประชาชน 42 คน

บทสรุป

สถิติเป็นศาสตร์ที่เกี่ยวข้องโดยตรงกับทุกคน เพราะการดำรงชีวิตของเราอยู่ที่การเปรียบเทียบ การวัด การประมาณค่าตลอดจนการนำตัวเลขมาเป็นเกณฑ์มาตรฐานต่าง ๆ อีกทั้งสถิติยังเป็นเครื่องมือที่ใช้ในทางวิทยาศาสตร์ที่นักวิทยาศาสตร์ใช้สรุปผลการทดลองต่าง ๆ ได้อย่างมากมาย โดยการที่จะใช้การวิเคราะห์ทางสถิติได้อย่างมีคุณภาพนั้นผู้วิเคราะห์ควรต้องทราบก่อนว่าข้อมูลที่เรานำมาวิเคราะห์นั้นเป็นข้อมูลประเภทใดหรือเป็นข้อมูลชนิดใด และควรเลือกสุ่มขนาดตัวอย่างมาใช้ในการวิเคราะห์เท่าใดเพื่อให้การวิเคราะห์โดยใช้วิธีการวิเคราะห์ทางสถิติแม่นยำมากขึ้น