บทที่ 5
การใช้สถิติเพื่อการวิจัย

หลังจากที่เก็บรวบรวมข้อมูลและตรวจสอบความถูกต้องของข้อมูลแล้วผู้วิจัยควรจัดหมวดหมู่ของข้อมูลด้วยหลักวิธีการทางสถิติที่ได้กำหนดไว้ในงานวิจัยนั้น ๆ ซึ่งสถิติที่นำมาใช้ในกระบวนการทำวิจัย นับว่ามีความสำคัญมาก เช่น การแจกแจงประชากร การสำรวจข้อมูล และการวิเคราะห์ข้อมูล รวมทั้งการประเมินค่าต่าง ๆ ให้ออกมาเป็นตัวเลข นอกจากนั้นสถิติ ยังเป็นประโยชน์สำหรับการวิเคราะห์เหตุการณ์ต่าง ๆ หรือวิเคราะห์ข้อมูลเพื่อทดสอบสมมติฐานในการวิจัยธุรกิจและวิจัยทางการบริหารตลอดจนการศึกษาวิจัยทางสังคมศาสตร์อื่น ๆ

ความหมายของสถิติ

สถิติ หมายถึง วิชาว่าด้วยการจัดเก็บข้อมูล การนำเสนอข้อมูลและการวิเคราะห์ข้อมูลและการคำนวณในการวิเคราะห์ข้อมูลมาสรุปผลเพื่อช่วยในการตัดสินใจในเหตุการณ์ได้อย่าง ถูกต้องและรวดเร็วยิ่งขึ้น ดังนั้น ข้อมูลที่ได้จากการวิเคราะห์ทางสถิติจะช่วยให้ผู้บริหารองค์การต่าง ๆ สามารถตัดสินใจได้อย่างถูกต้อง เช่น การคาดคะเนความต้องการของคู่แข่งขัน และส่วนแบ่งตลาดเป็นต้น ซึ่งมีการแบ่งวิชาสถิติออกเป็น 2 ส่วน คือ สถิติพรรณนา และสถิติอนุมาน

1. **สถิติพรรณนา** (Descriptive Statistics)

เป็นวิธีการรวบรวมข้อมูลเพื่อให้ได้ข้อมูลที่ถูกต้องจากแหล่งกำเนิดของข้อมูลหรือแหล่งที่มีผู้เก็บรวบรวมไว้ และเป็นการอธิบายลักษณะข้อมูลแต่ไม่ได้อ้างอิงไปดึงข้อมูลกลุ่มอื่น ๆ

2. **สถิติอนุมาน (**อ้างอิง**)** (Inference Statistics)

เป็นสถิติที่ใช้วิเคราะห์ข้อมูลที่เก็บรวบรวมจากตัวอย่าง เพื่อสรุปผลไปยังประชากรทั้งหมด ส่วนการอนุมานหรือการอ้างอิงจะถูกต้องมากน้อยเพียงใดขึ้นอยู่กับข้อมูลตัวอย่างที่เก็บรวบรวมมาว่าเป็นตัวแทนหรือตัวอย่างที่ดีเพียงใด ซึ่งจะต้องอาศัยการสุ่มตัวอย่างที่เหมาะสมด้วย

การนำสถิติมาใช้ในงานวิจัย

การนำหลักการและทฤษฎีสถิติมาประยุกต์ใช้กับกระบวนการวิจัย ซึ่งทำให้การทำวิจัยแต่ละขั้นตอนลดเวลาลงได้มาก โดยเฉพาะการคำนวณเพื่อวิเคราะห์ข้อมูลด้วย โปรแกรมทางคอมพิวเตอร์ทำให้ผลการคำนวณมีความถูกต้องและรวดเร็วยิ่งขึ้น

ประชากรและตัวอย่าง

การทำวิจัยแต่ละครั้งได้นำวิชาสถิติมาช่วยในการจำแนกตัวอย่างหรือกลุ่มตัวอย่างออกจากประชากร คือ

1. **ประชากร** (Population) เป็นหน่วยที่ผู้วิจัยมีความสนใจที่จะศึกษาฉะนั้นประชากรจึงหมายถึงบุคคล องค์กร สัตว์ สิ่งของต่าง ๆ ที่นำมาเป็นหน่วยศึกษาหรือปัญหาการวิจัย

2. **ตัวอย่าง** (Sample) เป็นส่วนหนึ่งของประชากร ซึ่งอาจเป็น 5%, 10% หรือ 50% ของจำนวนประชากรทั้งหมด โดยทั่วไปการทำวิจัยมักจะใช้ตัวอย่างหรือข้อมูลบางส่วน เพราะถ้าเก็บข้อมูลจากทุกหน่วยในประชากร จะทำให้เสียทั้งเวลาและค่าใช้จ่ายสูง ทำให้ไม่ทันเวลาที่กำหนดไว้ด้วย

3. **การสุ่มตัวอย่าง** (Random sampling) เป็นวิธีการเลือกตัวอย่างหรือการกระทำเพื่อให้ได้ตัวแทนประชากรให้มากที่สุดเท่าที่กระทำได้ โดยการสุ่มตัวอย่างมีหลักใหญ่ ๆ คือ

3.1 การสุ่มแบบไม่เป็นไปตามความน่าจะเป็น (Non Probability Sampling) เช่น การสุ่มแบบสะดวก (Convenience Sampling) การส่งแบบบังเอิญ (Accidental sampling) การสุ่มแบบโควตา (Quota sampling) และการสุ่มแบบเจาะจง (Purposive random sampling)

3.2 การสุ่มแบบให้เป็นไปตามโอกาสทางสถิติ (Probability Sampling) เช่น การสุ่มตัวอย่างแบบง่าย (Simple random sampling) การสุ่มตัวอย่างแบบมีระบบ (Systematic random sampling) การสุ่มตัวอย่างแบบกลุ่ม (Cluster random sampling) การสุ่มตัวอย่างแบบแบ่งชั้น (Stratify random sampling) การสุ่มตัวอย่างแบบหลายขั้น (Multi-stage random sampling) เป็นต้น

ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่เก็บรวบรวมมาวิเคราะห์มีทั้งข้อมูลปฐมภูมิและข้อมูลทุติยภูมิ ดังนี้

1. ข้อมูลปฐมภูมิ (Primary Data) เป็นข้อมูลที่ผู้วิจัยได้เก็บรวบรวมข้อมูลมาเอง เช่น การสอบถาม การสัมภาษณ์ การสังเกตการณ์ หรือการทดลอง โดยข้อมูลปฐมภูมิจะมีรายละเอียดตรงตามความต้องการ แต่ก็จะทำให้เสียเวลาและค่าใช้จ่ายสูง ซึ่งเป็นข้อมูลดิบที่ไม่ได้ทำการวิเคราะห์

2. ข้อมูลทุติยภูมิ (Secondary Data) เป็นข้อมูลที่ผู้วิจัยไม่ได้ทำการเก็บรวบรวมข้อมูลเอง แต่มีผู้อื่นหรือมีหน่วยงานได้รวบรวมข้อมูลไว้แล้ว เมื่อผู้วิจัยนำข้อมูลทุติยภูมิมาใช้จะต้องตรวจสอบข้อมูลอย่างละเอียดและระมัดระวังในการนำมาใช้ด้วย เพราะข้อมูลอาจคลาดเคลื่อนหรือผิดพลาดได้ ซึ่งจะมีผลทำให้การสรุปผิดพลาดไปด้วย ดังนั้น การทำวิจัยในแต่ละเรื่อง จึงมีการใช้ข้อมูลทั้ง 2 อย่าง ผสมกันไป

สำหรับข้อมูลทุติยภูมิส่วนใหญ่จะมีลักษณะ ดังนี้

2.1 ข้อมูลเชิงปริมาณ (Quantitative Data) เป็นข้อมูลที่สามารถวัดค่าออกมาเป็นตัวเลขได้ เช่น อายุ รายได้ ปริมาณสินค้า ยอดขายสินค้า ฯลฯ

2.2 ข้อมูลแบบต่อเนื่อง (Continuous Data) เป็นข้อมูลที่มีค่าได้ทุกค่าในช่วงที่กำหนด เช่น นำหนักสินค้า ความยาวเหล็ก ส่วนสูงของคน ความลึกของบ่อ รายได้เพิ่มขึ้น ฯลฯ

2.3 ข้อมูลแบบไม่ต่อเนื่อง (Discrete Data) เป็นข้อมูลที่มีค่าจำนวนเต็มหรือจำนวนนับได้ เช่น จำนวนคน จำนวนหน่วยงาน จำนวนห้องแถว ฯลฯ

2.4 ข้อมูลเชิงคุณภาพ (Qualitative Data) เป็นข้อมูลที่ไม่สามารถวัดค่าหรือคำนวณเฉลี่ยได้ เช่น เพศ การศึกษา ลักษณะของสินค้า หรือสีของเสื้อผ้า ฯลฯ

ตัวอย่างการนำสถิติมาใช้ในงานวิจัย

การนำสถิติมาวิเคราะห์ข้อมูลในงานวิจัย จะมีทั้งสถิติพรรณนา (Descriptive Statistics) และสถิติอนุมาน (อ้างอิง) Inference Statistics โดยมีวิธีการใช้ดังนี้

1. การใช้สถิติอธิบายลักษณะข้อมูล (Describing Data) เป็นการใช้สถิติมาวิเคราะห์ข้อมูลและอธิบายลักษณะของข้อมูลที่ใช้สถิติจำแนกออกเป็นค่าร้อยละ การแจกแจงออกเป็นความถี่ ($\overline{X}$) การคำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน (S.D.) ซึ่งมีทั้งข้อมูลส่วนบุคคล ข้อมูลเกี่ยวกับพฤติกรรม เช่น การเลือกตอบ (Check List) เป็นต้น

2. การใช้สถิติวิเคราะห์เพื่อสรุปจากข้อมูลตัวอย่าง (Conclusions from sample data) เป็นการใช้สถิติวิเคราะห์หาข้อสรุปเพื่อนำไปสู่การตอบวัตถุประสงค์การวิจัยโดยใช้สถิติเชิงอนุมานวิเคราะห์เพื่อนำไปสู่ข้อสรุปไปยังประชากร ซึ่งมีเทคนิคทางสถิติ ดังนี้

2.1 การใช้สถิติ Chi-Square เพื่อวิเคราะห์ข้อมูลที่อยู่ในรูปของความถี่ (frequency) ซึ่งเป็นข้อมูลเชิงคุณภาพหรือคุณลักษณะที่จำแนกข้อมูลเป็นประเภท เช่น จำนวนหรือความถี่ของแต่ละระดับ/แต่ละกลุ่ม เช่น ความคิดเห็นต่อสินค้า/ต่อบริหารจัดการ ซึ่งอาจแบ่งออกเป็น 4-5 ระดับ โดยใช้สถิติ χ^2 ทดสอบสมมติฐานสำหรับข้อมูลที่จำแนกทางเดียว เช่น การทดสอบลักษณะต่าง ๆ

และการแจกแจงของประชากร/ตัวอย่าง ว่าเป็นไปตามที่คาดหวังไว้หรือไม่ และทดสอบสมมติฐานสำหรับข้อมูลจำแนกแบบสองทาง ซึ่งเป็นการทดสอบความเป็นอิสระกันระหว่างข้อมูล 2 ลักษณะ (Testing for Independence) เช่น ปัจจัยส่วนบุคคลสัมพันธ์กับพฤติกรรมการซื้อหรือใช้บริการฯ

2.2 การใช้สถิติ t – test วิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบค่าเฉลี่ย ($\overline{X}$) โดยใช้กับข้อมูล 2 กลุ่ม และมีตัวแปรอิสระที่เป็นข้อมูลเชิงคุณภาพ/คุณลักษณะ และตัวแปรตามที่มีข้อมูลเป็นเชิงปริมาณ (คำนวณค่าเฉลี่ยได้) โดยใช้สถิติ t - test ได้สองรูปแบบ คือ กรณีตัวอย่าง 2 กลุ่มมีความสัมพันธ์หรือกลุ่มเดียวกัน (ไม่เป็นอิสระต่อกัน) ใช้ t-test Pairs และกรณีที่ตัวอย่าง 2 กลุ่มเป็นอิสระต่อกัน (ไม่เกี่ยวข้องกัน) ใช้ Independent sample t-test เช่น เพศชายและเพศหญิง โดยทดสอบเพื่อเปรียบเทียบ (ค่าเฉลี่ย) ระดับความพึงพอใจในการเลือกซื้อสินค้าหรือใช้บริการฯ

2.3 การใช้สถิติ F – test วิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบค่าเฉลี่ย ($\overline{X}$) โดยใช้กับข้อมูล 3 กลุ่มขึ้นไป มีทั้งการทดสอบความแปรปรวนภายในกลุ่มและความแปรปรวนระหว่างกลุ่ม และการทดสอบแบบ One-way ANOVA and Two-way ANOVA ถ้าการทดสอบตัวแปรตาม (ค่าเฉลี่ย) ปรากฏว่า “แตกต่างกันอย่างมีนัยสำคัญทางสถิติ” จะต้องนำไปเปรียบเทียบเป็นรายคู่ ๆ โดยวิธีของ Scheffe’s หรือ Fisher’s LSD. ตามความสอดคล้องต่อไป

2.4 การวิเคราะห์สหสัมพันธ์ (Multiple Correlation) เป็นการหาความสัมพันธ์ระหว่างตัวแปรตาม ที่ข้อมูลคำนวณค่าเฉลี่ยได้ทั้ง 2 ตัวแปร โดยคำนวณด้วย 4 วิธี คือ Product-Moment correlation, Phi’s correlation, Spearman’s rank and Kendall’s และเลือกใช้ตามเงื่อนไขฯ

การนำสูตรทางสถิติมาแสดงไว้โดยย่อ

เนื่องจากการเขียนเนื้อหาในบทที่เกี่ยวกับวิธีดำเนินการวิจัย เพื่อความเหมาะสมชัดเจน ผู้วิจัยควรนำสูตรทางสถิติที่ใช้วิเคราะห์ข้อมูลมาแสดงไว้ในบทนี้ด้วย เพื่อให้เห็นว่าผู้วิจัยจะใช้สถิติได้อย่างถูกต้อง ดังนั้น การนำสูตรคำนวณทางสถิติมาแสดงไว้ ควรสรุปย่อ ๆ ดังตัวอย่างต่อไปนี้

ตัวอย่าง การใช้สถิติวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ ได้ใช้สถิติเพื่อวิเคราะห์ข้อมูล ได้แก่ สถิติร้อยละ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน การวัดค่าความแปรปรวน การทดสอบค่าเฉลี่ยด้วย t-test and One way ANOVA และการวัดความสัมพันธ์ (อภินันท์ จันตะนี, 2549, หน้า 90-92)

1. การหาค่าร้อยละ (Percentage) โดยใช้สูตรดังนี้

$$P = \frac{f}{n} \times 100$$

f	หมายถึง	ความถี่หรือจำนวนข้อมูล
X	หมายถึง	ค่าของข้อมูลหรือคะแนน
n	หมายถึง	จำนวนตัวอย่างหรือผู้ตอบแบบสอบถาม
P	หมายถึง	ค่าร้อยละ (percentage)

2. ค่าเฉลี่ยตัวอย่าง (Sample mean = $\overline{X}$) โดยใช้สูตรดังนี้

$$\overline{X} = \frac{\sum fx}{n}$$

เมื่อ	$\overline{X}$	หมายถึง	ค่าเฉลี่ยของตัวอย่าง
	$\sum$fX	หมายถึง	ผลรวมของข้อมูลทั้งหมด
	n	หมายถึง	จำนวนตัวอย่างที่ตอบแบบสอบถาม

3. ส่วนเบี่ยงเบนมาตรฐาน (**S.D.**) โดยใช้สูตรดังนี้

$$S.D. = \sqrt{\frac{n(\sum fx^2) - (\sum x)^2}{n(n-1)}}$$

เมื่อ	**S.D.**	หมายถึง	ส่วนเบี่ยงเบนมาตรฐาน
	$\sum fx^2$	หมายถึง	ผลรวมกำลังสอง
	$(\sum fx)^2$	หมายถึง	ผลรวมของข้อมูล
	N	หมายถึง	ขนาดของตัวอย่าง

เมื่อรวบรวมข้อมูลและแจกแจงความถี่แล้ว จะใช้คะแนนเฉลี่ยของกลุ่มตัวอย่างมาพิจารณาระดับความพึงพอใจ ซึ่งมีเกณฑ์ในการพิจารณา (ล้วน และอังคณา สายยศ, 2536, หน้า 157)

ระดับ = $\frac{\text{คะแนนสูงสุด-คะแนนต่ำสุด}}{\text{จำนวนชั้น}}$

= $\frac{5-1}{5}$ = **0.8**

ค่าเฉลี่ย 4.20 - 5.00 หมายถึง ระดับความพึงพอใจมากที่สุด

ค่าเฉลี่ย 3.40 - 4.19 หมายถึง ระดับความพึงพอใจมาก

ค่าเฉลี่ย 2.60 - 3.39 หมายถึง ระดับความพึงพอใจปานกลาง

ค่าเฉลี่ย 1.80 - 2.59 หมายถึง ระดับความพึงพอใจน้อย

ค่าเฉลี่ย 1.00 - 1.79 หมายถึง ระดับความพึงพอใจน้อยที่สุด

4. การวิเคราะห์ความแปรปรวนกรณีตัวอย่างอิสระกัน และประชากร 2 กลุ่มไม่เท่ากัน

เป็นการเปรียบเทียบระดับความคิดเห็นของตัวอย่าง 2 กลุ่มที่เป็นอิสระกัน **โดยใช้สูตร t - test** เพื่อเปรียบเทียบค่าเฉลี่ยของแต่ละคู่ ดังนี้

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

5. การวิเคราะห์ความแปรปรวนกรณีตัวอย่างประชากร 3 กลุ่มขึ้นไป

เป็นการเปรียบเทียบระดับความคิดเห็นของตัวอย่าง ตั้งแต่ 3 กลุ่มขึ้นไปโดยใช้สถิติ **F – test** เพื่อทดสอบค่าเฉลี่ย มีสูตรดังนี้

$$F = \frac{MS_B}{MS_W}$$

เมื่อทดสอบด้วยสถิติ F – test ปรากฏว่าแตกต่าง (**Sig.**) ให้แยกเปรียบเทียบเป็นรายคู่ **(Post Hoc** Comparison) ด้วยวิธีของ **Scheffe's** และ Fisher's **LSD** มีสูตรคำนวณ ดังนี้

1) **วิธีของ Scheffe** มีสูตรดังนี้

$$F = \frac{(\overline{X}_1 - \overline{X}_2)^2}{MSW\left(\frac{1}{n_1} + \frac{1}{n_2}\right)(k-1)}$$

2) **วิธีของ Fisher's LSD** มีสูตรดังนี้

$$LSD = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)(MSW)F}$$

6. การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ที่มี 2 กลุ่มและเป็นค่าความถี่ฯ

เป็นการหาความสัมพันธ์ระหว่าง ตัวแปรที่เป็นปัจจัยส่วนบุคคล กับ ตัวแปรตามที่คำนวณค่าออกมาเป็นความถี่/ร้อยละทั้ง 2 กลุ่ม โดยใช้สถิติ Chi-Square มีสูตรคำนวณ ดังนี้

1) การทดสอบสมมติฐานแบบ Goodness of Fit Test ที่ใช้ทดสอบข้อมูลทางเดียว มีสูตรคำนวณ ดังนี้

$$\chi^2 = \sum_{i=1}^{n} \frac{(oi - Ei)^2}{Ei}$$

กำหนดให้ χ^2 แทนค่าไค-สแควร์ (Chi-Square)

O ค่าความถี่ที่ได้จากการสังเกต (Observed Frequency)

E ค่าความถี่ที่ได้จากความคาดหวัง (Expert Frequency)

N จำนวนกลุ่มตัวแปร และกรณี df = n – 1

E = **N**/(กลุ่ม/ระดับ/ค่าเฉลี่ย)

2) การทดสอบความเป็นอิสระกัน (χ^2 testing for Independence) โดยใช้กับกรณีที่ข้อมูล **2 x 3, 3 x 3, 3 x 5 or 5 x 4** ใช้สูตร ดังนี้

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

กำหนดให้ χ^2 **แทนค่าไค-สแควร์** (Chi-Square)

O_{ij} ค่าความถี่ที่ได้จากการสังเกต (Observed Frequency)

E_{ij} ค่าความถี่ที่ได้จากความคาดหวัง (Expert Frequency)

n จำนวนกลุ่มตัวแปร และกรณี df = n – 1

r_{ij} จำนวนข้อมูลคุณลักษณะตามแนวนอนในระดับที่ i….

c_{ij} จำนวนข้อมูลคุณลักษณะตามแนวตั้งในระดับที่ j….

E = **R** x **C** / **N**

7. การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ที่มี 2 กลุ่มและเป็นค่าเฉลี่ยด้วย เป็นการหาความสัมพันธ์ระหว่าง ตัวแปรที่เป็นตัวแปรตาม โดยวัดความสัมพันธ์ระหว่างตัวแปรตามเหมือนกัน แต่มีความต่างกลุ่มกันและค่าคำนวณฯได้เป็นค่าเฉลี่ยทั้ง 2 กลุ่ม โดยใช้สถิติ Spearman Rank Correlation วิเคราะห์ มีสูตรคำนวณ ดังนี้

$$r = \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{\sum x^2 - (\sum x)^2/n}\sqrt{\sum y^2 - (\sum y)^2/n}}$$

เมื่อ **x, y** หมายถึง ค่าที่สามารถคำนวณได้จากข้อมูล

n หมายถึง จำนวนข้อมูลของแต่ละตัวแปร

สำหรับการแปลความหมายค่าสหสัมพันธ์ (Multiple Correlation) โดยมีเกณฑ์วัดระดับความสัมพันธ์ (อภินันท์ จันตะนี, 2549, หน้า 7) ดังนี้

- ค่าสหสัมพันธ์ .01 - .20 มีความสัมพันธ์กันในระดับต่ำมาก
- ค่าสหสัมพันธ์ .21 - .40 มีความสัมพันธ์กันในระดับต่ำ
- ค่าสหสัมพันธ์ .41 - .60 มีความสัมพันธ์กันในระดับปานกลาง
- ค่าสหสัมพันธ์ .61 - .75 มีความสัมพันธ์กันในระดับค่อนข้างสูง
- ค่าสหสัมพันธ์ .76 - .90 มีความสัมพันธ์กันอยู่ในระดับสูง
- ค่าสหสัมพันธ์ .91 – 1.00 มีความสัมพันธ์กันในระดับสูงมาก

การใช้สถิติเพื่อแจกแจงประชากร

ประชากรในทางสถิติ ได้หมายถึง คน สัตว์ สิ่งของ องค์กร หรือสถานที่ต่าง ๆ ที่ผู้วิจัยศึกษาส่วนตัวอย่างเป็นส่วนหนึ่งของประชากร เช่น 5 – 10% หรือ 20% ทั้งนี้ ขึ้นอยู่กับความเหมาะสมหรือการเป็นตัวแทนหรือตัวอย่างที่ดีของประชากรที่ใช้ในการวิจัย แต่การที่จะใช้ตัวอย่าง (Sample Size) เท่าใดจึงจะเป็นตัวแทนที่ดีนั้น ในทางสถิติได้ทดลองและมีความเชื่อมั่น ดังต่อไปนี้

การกำหนดขนาดตัวอย่าง (Sample Size)

การสำรวจความคิดเห็นหรือทัศนคติจากตัวอย่าง โดยทั่วไปย่อมมีความคลาดเคลื่อนอยู่บ้าง ถ้าไม่ต้องการให้เกิดความคลาดเคลื่อนก็ต้องใช้ประชากรทั้งหมด แต่การที่จะยอมให้เกิดความ

คลาดเคลื่อนเท่าใดนั้น นับว่าผู้วิจัยจะต้องให้ความสำคัญ เพราะถ้ายอมให้คลาดเคลื่อนสูง ก็จะทำให้ผลการวิจัยมีความน่าเชื่อถือน้อยหรือด้อยคุณภาพ ดังนั้น การทำวิจัยจึงจำเป็นต้องนำวิชาสถิติมาช่วยเพื่อกำหนดค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยตัวอย่าง (σ X) ซึ่งเป็นค่าที่วัดการกระจายของ X รอบ ๆ ค่าเฉลี่ย (μ) ของประชากร (σ X) มีค่าน้อยจะทำให้ค่าเฉลี่ย X ของตัวอย่างเข้าใกล้ค่า μ (ค่าเฉลี่ยประชากร) มากขึ้นไปอีกโดยที่ σ = ส่วนเบี่ยงเบนมาตรฐานของประชากร

อย่างไรก็ตามการที่ค่า σ X จะลดลงมากน้อยเพียงใดขึ้นอยู่กับค่า n เพราะ n เป็นตัวหาร เช่น $\sigma\overline{X} = \sigma x / \sqrt{n}$

1. การกำหนดตัวอย่างที่ทำให้เกิดความคลาดเคลื่อน (เชื่อมั่นลดลง) **ต่างกัน**

ตัวอย่าง n = 10, σ = 100 จะได้ $\sigma\overline{X}$ = $100/\sqrt{10}$ = 31.63%

n = 30, σ = 100 จะได้ $\sigma\overline{X}$ = $100/\sqrt{30}$ = 18.26%

n = 50, σ = 100 จะได้ $\sigma\overline{X}$ = $100/\sqrt{50}$ = 14.14%

n = 100, σ = 100 จะได้ $\sigma\overline{X}$ = $100/\sqrt{100}$ = 10%

n = 200, σ = 100 จะได้ $\sigma\overline{X}$ = $100/\sqrt{200}$ = 7.07%

n = 400, σ = 100 จะได้ $\sigma\overline{X}$ = $100/\sqrt{400}$ = 5%

2. การกำหนดตัวอย่างที่ทำให้เกิดความคลาดเคลื่อนมาตรฐานที่ 5%

การคำนวณหาจำนวน/ขนาดของตัวอย่าง กรณีที่ประชากรมีขนาดใหญ่และไม่ทราบจำนวนประชากรที่แน่นอน ถ้าทราบจำนวนประชากร สามารถคำนวณด้วยสูตรของ **Taro Yamane** โดยกรณีไม่มีการบันทึกข้อมูลจำนวนประชากรไว้อย่างชัดเจน สามารถใช้สูตรเพื่อคำนวณหาขนาดของกลุ่มตัวอย่าง ได้ดังนี้

1) กรณีที่ผู้วิจัย ไม่สามารถหาข้อมูลเกี่ยวกับจำนวนประชากรที่จะนำมาใช้ในการวิจัยไว้อย่างชัดเจน (Infinite population) จึงใช้สูตรเพื่อคำนวณตัวอย่าง (อภินันท์ จันตะนี, 2549, หน้า 35)

สูตร $$n = \frac{3.84(d)^2}{(S)^2}$$

ตัวอย่างที่ 1 โดยต้องการกลุ่มตัวอย่างเพื่อใช้ในการทำวิจัยหรือวิทยานิพนธ์ โดยยอมให้เกิดความคลาดเคลื่อนขึ้นได้ 5% (0.05)

วิธีทำให้ S = 1.96 (สำหรับความเชื่อมั่น 95%)

d = 0.05 (คลาดเคลื่อนได้ 5% หรือ = 0.05)

โดยแทนค่า S ที่ระดับนัยสำคัญทางสถิติมีค่าเท่ากับ 1.96 (S = 1.96)

และแทนความคลาดเคลื่อน (d = .05) ที่คาดว่าจะยอมให้เกิดขึ้นได้ (d)

$$n = \frac{3.84(.05)^2}{(1.96)^2}$$

$$\mathbf{n} = \frac{\mathbf{1}}{0.002499} = \mathbf{400}.16$$

ดังนั้น จึงใช้จำนวน**กลุ่มตัวอย่าง 400 ราย**

ตัวอย่างที่ 2 โดยต้องการตัวอย่างไม่มากนักและใช้วิธีเก็บข้อมูลเชิงลึกคือใช้ได้ทั้งวิธีเชิงปริมาณและเชิงคุณภาพหรือต้องการใช้ตัวอย่างสำหรับทำเป็นระดับภาคนิพนธ์หรือการค้นคว้าอิสระ (IS) โดยยอมให้เกิดความคลาดเคลื่อนขึ้นได้ 7% (0.07)

วิธีทำ S = 1.96 (สำหรับความเชื่อมั่น 95%)

d = 0.07 (คลาดเคลื่อนได้ 7% หรือ = 0.07)

$$n = \frac{3.84(.07)^2}{(1.96)^2}$$

$$\mathbf{n} = \frac{\mathbf{1}}{0.0048979} = \mathbf{204.17}$$

โดยได้จำนวน**กลุ่มตัวอย่าง 204 ราย**

ตัวอย่างที่ 3 กรณีที่ต้องการใช้ตัวอย่างเพื่อฝึกหัดทำวิจัยหรือสำหรับทำวิจัยเพื่อใช้เป็นภาคนิพนธ์ (Term-paper) หรือ**การค้นคว้าอิสระ (IS)** ที่มี **3 หน่วยกิ** โดยยอมให้เกิดความคลาดเคลื่อนขึ้นได้เท่ากับ 10% (0.10)

วิธีทำ S = 1.96 (สำหรับความเชื่อมั่น 95%)

d = 0.10 (คลาดเคลื่อนได้ 10% หรือ = 0.10)

$$n = \frac{3.84(.10)^2}{(1.96)^2}$$

$$\mathbf{n} = \frac{\mathbf{1}}{0.0099958} = \mathbf{100.}04$$

โดยได้จำนวน**ตัวอย่าง 100 ราย**

อย่างไรก็ตาม จำนวนกลุ่มตัวอย่างหรือกลุ่มตัวอย่างที่คำนวณได้ในแต่ละสูตรเป็นเพียงจำนวนตัวอย่างเบื้องต้นเท่านั้น แต่ที่สำคัญไปกว่านั้น คือ วิธีการสุ่มตัวอย่างเพื่อให้ได้ตัวแทนของประชากรที่แท้จริงถือว่าสำคัญที่สุด ทั้งนี้ เพราะถ้าวิธีการสุ่มที่จะได้ตัวแทนตัวอย่างที่ดีแล้วก็ไม่จำเป็นจะต้องใช้ตัวอย่างมาก ๆ ตามที่คำนวณได้ในแต่ละสูตร ดังนั้นวิธีการสุ่มตัวอย่างเพื่อให้ได้ตัวแทนที่ดี จึงมีความสำคัญด้วย

2) การกำหนดขนาดตัวอย่างที่มีการบันทึกข้อมูลจำนวน/ปริมาณประชากรไว้

การคำนวณหาขนาด/จำนวนตัวอย่างในกรณีที่สามารถรู้จำนวนหรือปริมาณประชากร เช่น **จำนวนผู้บริโภค ผู้ผลิต ยอดขาย**หรือ**รายได้** ฯลฯ เช่น การหา**ขนาดตัวอย่าง** (n) ได้ 2 วิธีดังนี้

2.1) การคำนวณหาขนาดกลุ่มตัวอย่างกรณีที่ทราบจำนวนประชากรชัดเจน ถ้าประชากรมีจำนวน 1,500 ราย และให้ความเชื่อมั่น 95% โดยใช้สูตรของ Taro Yamane (1973) ดังนี้

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

วิธีทำ

$$n = \frac{1{,}500}{1+1{,}500(.05)^2}$$

$$= \frac{1{,}500}{1+3.75} = \frac{1{,}500}{4.5}$$

= 315.79 ราย

อย่างไรก็ตาม ตัวอย่างที่คำนวณได้ 316 นี้ เป็นจำนวนตัวอย่างขั้นต่ำสุดเท่านั้น แต่ถ้าต้องการให้เกิดความเชื่อมั่นเพิ่มขึ้น สามารถเพิ่มเป็น 320 ราย หรือ 400 รายก็ได้ แต่จะต้องมีหลักการหรือเหตุผลและอธิบายไว้ให้ชัดเจน ทั้งนี้ เพื่อลดความคลาดเคลื่อนจากกลุ่มตัวอย่างลงนั่นเอง และมีข้อพึงระวัง คือ จำนวนตัวอย่างเมื่อจำแนกออกเป็นกลุ่ม ๆ แต่ละกลุ่มไม่ควรต่ำกว่า 30 ราย หรือ 50 ราย เพราะจะทำให้สามารถใช้สถิติเพื่อได้เป็นอย่างดีหรือสอดคล้องฯ

2.2) สำหรับกลุ่มตัวอย่างที่ใช้ในกรณี “การค้นคว้าอิสระ (IS) หรือให้นักศึกษาฝึกปฏิบัติทำวิจัยฯ” โดยสามารถใช้ขนาดของกลุ่มตัวอย่างลดลงเป็นกึ่งหนึ่งจากสูตรของ Taro Yamane (1973) (ที่เชื่อมั่น 95% และคลาดเคลื่อน 5%) โดยสูตรดังกล่าวมีเงื่อนคือให้ความเชื่อถือได้ 93% และให้มีความคลาดเคลื่อน 7% (d = .07) (อภินันท์ จันตะนี, 2549, หน้า 29) ถ้ามีประชากร 1,500 รายหาได้ดังนี้

$$n = \frac{N}{2+N(d)^2}$$

วิธีทำ

$$n = \frac{1500}{2+1500(.07)^2}$$

$$= \frac{1{,}500}{2+7.35} = \frac{1{,}500}{9.35}$$

= 160.43 ราย

สำหรับประชากรที่มีจำนวน 20,000 ราย เมื่อใช้สูตรดังกล่าว จะได้กลุ่มตัวอย่างเท่ากับ 200 ราย ดังตัวอย่าง โดยให้ความเชื่อมั่น 93% และมีความคลาดเคลื่อน 7% (d = .07) ซึ่งสามารถคำนวณได้ (อภินันท์ จันตะนี, 2549, หน้า 29) ดังนี้

$$n = \frac{N}{2 + N(d)^2}$$

วิธีทำ

$$n = \frac{20{,}000}{2 + 20{,}000(.07)^2}$$

$$= \frac{20{,}000}{2 + 98} = \frac{20{,}000}{100}$$

= 200 ราย

โดยการใช้สูตรดังกล่าว มีเงื่อนไข คือ ต้องใช้กับประชากรหรือตัวอย่างที่เป็นลูกค้าประจำหรือเป็นสมาชิก หรือเป็นพนักงานในสถานประกอบการนั้น ๆ จึงจะสามารถนำสูตรไปใช้ได้อย่างเหมาะสม และเป็นกรณีที่จะสามารถสุ่มตัวอย่างที่เป็นตัวแทนของประชากรได้

การทดลองเครื่องมือหรือแบบสอบถาม

เมื่อสร้างเครื่องมือเสร็จแล้วก็ส่งไปให้ผู้ทรงคุณวุฒิหรือกรรมการที่ปรึกษาซึ่งจะตรวจว่าตรงตามเนื้อหา/ตรงตามโครงสร้างหรือสอดคล้องกับวัตถุประสงค์และกรอบแนวคิดของการวิจัยนั่นเอง ถ้ากรรมการ/ผู้ทรงคุณวุฒิเห็นชอบแล้ว จึงจะนำเครื่องมือไป**ทดลอง** (Try-out) กับประชากร/ตัวอย่าง ที่ไม่ใช่กลุ่มตัวอย่างที่จะสำรวจหรือเก็บข้อมูลจริง โดยใช้ทดลองไม่ต่ำกว่า 30 ราย เมื่อเก็บรวบรวมข้อมูลได้ครบจำนวน 30 รายแล้ว ก็นำมาวิเคราะห์เพื่อหาความเชื่อมั่นด้วยค่า Alpha (α) ของ Cronbach's Alpha (1974)

โดยมีสูตรที่ใช้ในการคำนวณค่าสัมประสิทธิ์อัลฟา (α - Coefficient) มีดังนี้

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum V_i}{V_t}\right]$$

เมื่อ α = ค่าระดับความเชื่อถือได้

k = จำนวนข้อมูล

V_i = ความแปรปรวนของข้อมูลแต่ละข้อ

V_t = ความแปรปรวนของข้อมูลรวมทุกข้อ

เมื่อได้ทดลอง (Try-Out) เพื่อหาค่าความเชื่อมั่นจากแบบสอบถาม โดยการคำนวณ Alpha Coefficient ได้แล้ว (มีค่าเฉลี่ยรวมไม่ต่ำกว่า .65 หรือ 65%) และให้นำมาอ้างอิงไว้ในบทที่ 3

การสุ่มตัวอย่าง (Random Sampling)

การทำวิจัยส่วนใหญ่ผู้วิจัยไม่สามารถสำรวจหรือสอบถามจากประชากรได้ทั้งหมด จึงจำเป็นต้องเลือกตัวแทนหรือตัวอย่างมาเพื่อสอบถามหรือสัมภาษณ์ แต่การเลือกตัวอย่างเพื่อทำวิจัยจะต้องให้ปราศจากความลำเอียง (Bias) ดังนั้น การเลือกตัวอย่างหรือการสุ่มตัวอย่าง (Random) เพื่อให้ได้ตัวแทนหรือตัวอย่างที่เหมาะสมและไม่ลำเอียง จึงนิยมเลือกตัวอย่าง 2 ตามหลักการ ดังนี้

1. การเลือกตัวอย่างแบบไม่เป็นไปตามความน่าจะเป็น (Non Probability Sampling)

เป็นการเลือกตัวอย่างที่ไม่ทราบโอกาสที่แต่ละหน่วยของประชากรที่จะถูกเลือก จึงไม่จำเป็นจะต้องทราบรายชื่อทุกหน่วยของประชากร ซึ่งวิธีนี้อาจไม่เหมาะสมสำหรับการวิจัยปัญหาบางเรื่อง โดยทั่วไปนิยมเลือกหรือสุ่มตัวอย่างอยู่ 5 วิธี ดังนี้

1) การสุ่มแบบตามความสะดวก (Convenience Sampling)

เป็นการเลือกตัวอย่างที่ไม่มีการยึดหลักการใด ๆ เพียงแต่เลือกหน่วยตัวอย่างตามความสะดวก เช่น การเลือกลูกค้าธนาคารตามชุมชนต่าง ๆ เพื่อสำรวจความคิดเห็น หรือการสอบถามประชาชนทั่วไปในกรุงเทพมหานคร ในเรื่องปัญหาจราจร ฯลฯ ซึ่งเป็นการสอบถาม ตามสถานที่ต่าง ๆ ที่สะดวก โดยสอบถามจนได้ตัวอย่างครบตามจำนวน (Sample Size) ที่กำหนดไว้

2) การสุ่มแบบบังเอิญ (Accidental sampling)

เป็นการเลือกตัวอย่างแบบไม่ได้ยึดตามหลักเกณฑ์ เพียงแต่ตั้งเป้าหมายของตัวอย่างให้ตรงกับวัตถุประสงค์การวิจัย ดังเช่น การสำรวจความคิดเห็นของนักเรียน/นักศึกษาต่อโภชนาการโรงอาหารในโรงเรียนหรือมหาวิทยาลัยฯ หรือการสำรวจความคิดเห็นของลูกค้าที่มาใช้บริการในธนาคารหรือร้านอาหารฯ ถ้าพบใคร (โดยบังเอิญ) สามารถสอบถามความคิดเห็นได้ทันที แต่ถ้าเป็นการสอบถามทุกคนที่เดินเข้ามาซื้อสินค้าในร้านหรือถามทุกคนที่เข้ามาใช้บริการในหน่วยงานหรือสถานประกอบกา/บริษัทฯ ไม่เป็นลักษณะการบังเอิญ (Accidental sampling) แต่อาจเป็น Purposive

3) การสุ่มแบบเจาะจง (Purposive sampling)

เป็นการเลือกตัวอย่างที่ใช้เหตุผลและวิจารณญาณในการเลือก เช่น การเลือกตัวอย่างจากผู้ที่คาดว่าจะเป็นตัวแทนหรือตัวอย่างที่ดีและสามารถตอบปัญหาต่าง ๆ แทนประชากรทั้งหมดได้ เช่น เลือกหัวหน้า/รองหัวหน้า/เลขาฯ หรือเลือกประธาน/เลขาฯ หรือหัวหน้ากลุ่ม/ชุมชนต่าง ๆ เพื่อเป็นตัวแทนมาสัมภาษณ์หรือตอบแบบสอบถาม และอาจเป็นตัวอย่างให้การสังเกต เป็นต้น

4) การสุ่มแบบโควต้า (Quota sampling)

เป็นการเลือกตัวอย่างที่ใช้หลักเกณฑ์ในการเลือก เช่น การกำหนดจำนวนตัวอย่างจากแต่ละกลุ่มที่เป็นสัดส่วนกับจำนวนประชากรแต่ละกลุ่มหรือแต่ละแผนกในองค์การ หรือฝ่ายต่าง ๆ ในบริษัท หรือกำหนดสัดส่วนของลูกค้า/ประชาชนในชุมชนจากจำนวนมาเป็นตัวอย่าง

5) การสุ่มแบบลูกโซ่ (Snowball Sampling)

เป็นลักษณะการเขียนจดหมายลูกโซ่ กล่าวคือ ถ้าผู้วิจัยได้เก็บข้อมูลหรือสัมภาษณ์บุคคลหนึ่งแล้ว ก็ให้บุคคลนั้นแนะนำบุคคลอื่นต่อ ๆ กันไปเรื่อย ๆ จนกระทั่งได้ตัวอย่างครบตามจำนวนเท่าที่กำหนดในขนาดตัวอย่าง (Sample size)

2. การเลือกหรือสุ่มตัวอย่างแบบให้เป็นไปตามโอกาสทางสถิติ (Probability Sampling)

เป็นการเลือก/สุ่มตัวอย่างที่สามารถเปิดโอกาสให้แต่ละหน่วยของประชากรจะถูกเลือกมาเป็นตัวแทนหรือตัวอย่างเท่า ๆ กัน โดยมี 5 วิธี ดังนี้

1) การสุ่มตัวอย่างแบบง่าย (Simple random sampling)

เป็นการสุ่มตัวอย่างที่เปิดโอกาสให้แต่ละหน่วยตัวอย่างมีโอกาสถูกเลือกมาเท่า ๆ กัน เพราะลักษณะของประชากรต้องมีการกระจาย การสุ่มที่ให้ได้เป็นตัวแทนของประชากรที่ดี เช่น ต้องการศึกษาทัศนคติต่อมหาวิทยาลัยฯ ของนักศึกษากลุ่มวิชาการจัดการทั่วไปจากความเชื่อว่านักศึกษากลุ่มวิชาการจัดการทั่วไป น่าจะมีทัศนคติต่อมหาวิทยาลัยฯ เหมือน ๆ กัน ดังนั้นการสุ่มตัวอย่างแบบง่าย จะสามารถทำได้ เพราะเพียงแต่ให้โอกาสในการสุ่มในแต่ละครั้ง จากนักศึกษาสาขาบริหารธุรกิจ กลุ่มวิชาการจัดการทั่วไป ให้มีโอกาสถูกสุ่มเท่าเทียมกัน โดยทั่วไปจะใช้วิธีจับฉลากหรือใช้กับกลุ่มผู้บริโภคหรือลูกค้าที่มาใช้บริการ

2) การสุ่มตัวอย่างแบบมีระบบ (Systematic random sampling)

เป็นกรณีกลุ่มประชากรที่จะทำการสุ่มได้ถูกจัดไว้เป็นระบบอยู่แล้ว เช่น เรียงตามเลขพนักงานฯ หรือเรียงลำดับตามบัญชีรายชื่อในการเลือกตั้ง หรือครัวเรือนตามบ้านเลขที่ สามารถจัดระบบ โดยนำทุก ๆ ลำดับที่ 3 หรือที่ 5 มาเป็นตัวอย่าง ซึ่งมีความเชื่อที่ว่าประชากรจะเรียงลำดับกันเป็นระบบอยู่แล้ว

3) การสุ่มตัวอย่างแบบกลุ่ม (Cluster random sampling)

เป็นการสุ่มตัวอย่างจากแต่ละกลุ่ม เพราะมีความเชื่อว่าแต่ละกลุ่มเป็นตัวแทนของประชากรอยู่แล้ว เช่น การสุ่มตัวอย่างเพื่อจะศึกษาการใช้คอมพิวเตอร์ในมหาวิทยาลัยฯ หรือลูกค้าที่ซื้อโทรศัพท์มือถือ หรือกลุ่มสมาชิกสหกรณ์ประเภทต่าง ๆ โดยแบ่งกลุ่มลูกค้าหรือสมาชิกหรือนักศึกษาออกไปตามชั้นปี และการแบ่งออกเป็นกลุ่ม ๆ เพื่อให้กระจายตัวอย่างออกไปอย่างทั่วถึง

ดังนั้น การแบ่งตัวอย่างออกเป็นกลุ่ม ๆ จึงเหมาะสำหรับการทำวิจัยที่เน้นลูกค้าหรือสมาชิกเป็นเป้าหมายสำคัญ

4) การสุ่มตัวอย่างแบบแบ่งชั้น (Stratify random sampling)

เป็นการแบ่งกลุ่มตัวอย่างโดยแบ่งออกเป็นชั้น ๆ (Strata) เพราะมีความเชื่อว่าประชากรมีความแตกต่างกันมากตามตัวแปรคุณลักษณะ เช่น เพศ ช่วงอายุ ระดับการศึกษา รายได้ อาชีพ ฯลฯ ดังนั้น การแยกตัวแปรอิสระต่าง ๆ ออกมาเป็นชั้น ๆ เพื่อกระจายให้ตัวอย่างที่ได้รับเลือกและมีโอกาสเป็นตัวแทนหรือตัวอย่างของทุกระดับชั้น ซึ่งจะทำให้เป็นตัวแทนหรือตัวอย่างที่ดีได้

5) การสุ่มตัวอย่างแบบหลายขั้น (Multi-stage random sampling)

เป็นการนำวิธีการสุ่มตัวอย่างทุกแบบมาผสมผสานกันโดยแบ่งการสุ่มตัวอย่างออกเป็นขั้นตอนต่าง ๆ เช่น การศึกษารูปแบบของธุรกิจชุมชนในท้องถิ่น ซึ่งมีขั้นตอน ดังนี้

ขั้นตอนที่ 1 การสุ่มตัวอย่างแบบแบ่งชั้น สุ่มจังหวัดในประเทศไทยมาจากชั้นที่เป็นภาคภูมิศาสตร์ ได้แก่ ภาคกลาง ภาคเหนือ ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ โดยเชื่อว่าเรื่องที่ต้องการศึกษาน่าจะมีรูปแบบในการพัฒนาแตกต่างกันไปตามตัวแปรของภูมิภาค จึงแบ่งชั้นเพื่อให้ได้กลุ่มตัวอย่างเป็นตัวแทนจากภาคต่าง ๆ

ขั้นตอนที่ 2 การสุ่มตัวอย่างแบบง่าย หลังจากได้จังหวัดที่เป็นตัวแทนจากทุกภาคของประเทศแล้ว ก็ทำการสุ่มจากอำเภอ โดยให้ทุกอำเภอภายในแต่ละจังหวัดตัวอย่างมีโอกาสถูกเลือกโดยเท่าเทียมกัน เพราะเชื่อว่าอำเภอในจังหวัดตัวอย่าง ก็เป็นตัวแทนของจังหวัดนั้น ๆ เท่ากัน

ขั้นตอนที่ 3 การสุ่มตัวอย่างแบบกลุ่ม เมื่อได้อำเภอเป็นตัวอย่างแล้ว ใช้อำเภอเป็นกลุ่ม (cluster) เพื่อกำหนดการเลือกตำบลออกมาเป็นตัวอย่าง โดยให้เป็นไปตามสัดส่วนแต่ละตำบลภายในอำเภอ ซึ่งวิธีการนี้ จะได้ตำบลที่เป็นตัวแทนจากอำเภอต่าง ๆ มีโอกาสได้รับเลือกเท่ากัน

ขั้นตอนที่ 4 การสุ่มตัวอย่างแบบมีระบบ เพื่อให้ได้ตัวแทนหมู่บ้านที่กระจายอยู่ในตำบล โดยใช้จำนวนหมู่บ้านมาเป็นการเลือกตัวอย่าง แต่เมื่อถึงประชาชนในหมู่บ้านอาจใช้วิธีสุ่มจากประชาชนทั้งหมดในหมู่บ้าน และการเลือกตัวอย่างไม่จำเป็นต้องใช้การสุ่มตัวอย่างเหมือนกันทุกขั้นตอน แต่ต้องผสมผสานกันหลายขั้นตอนได้ ทั้งนี้ ขึ้นอยู่กับการเป็นตัวแทนหรือตัวอย่างที่ดีของการทำวิจัยในแต่ละเรื่อง เป็นต้น

ข้อควรระวัง สำหรับการกำหนดขนาดตัวอย่าง (Sample Size) ที่เหมาะสมและการสุ่มตัวอย่าง (Random Sampling) โดยการสุ่มตัวอย่างจะต้องมีความสัมพันธ์กับขนาดของตัวอย่างและขนาดตัวอย่างจะต้องมีที่มาหรืออิงหลักการหรือทฤษฎีทางสถิติได้ด้วย

1. การสุ่มตัวอย่าง ถ้าวิธีการสุ่มตัวอย่างเป็นวิธีการที่พอจะเชื่อมั่นได้ว่าตัวอย่างที่ทำการสุ่มมาเป็นตัวแทนของประชากรได้จริง ๆ ขนาดตัวอย่างหรือจำนวนตัวอย่างก็ไม่จำเป็นจะต้องมีจำนวนหรือปริมาณมาก ๆ แต่ถ้าเก็บข้อมูลจากประชากรทั้งหมด ก็ไม่ต้องใช้วิธีการสุ่มตัวอย่าง

2. การกำหนดตัวอย่างที่เหมาะสม ถ้าหากกำหนดจำนวนตัวอย่างตามความคิดหรือความคาดหวังของผู้วิจัยเองแล้ว นับว่าเป็นเรื่องเชื่อได้ยาก โดยเฉพาะผู้ทำวิจัยเป็นครั้งแรก (มือใหม่) จึงจำเป็นจะต้องอ้างอิงทฤษฎีเกี่ยวกับขนาดของตัวอย่างหรือตัวแทน ดังเช่น ถ้าไม่ทราบจำนวนประชากรหรือลูกค้ามาซื้อสินค้าหรือใช้บริการ ซึ่งจำเป็นต้องใช้สูตรคำนวณหาจำนวนตัวอย่างและกำหนดความคลาดเคลื่อนไว้ด้วย แต่ถ้าทราบจำนวนประชากรที่แน่นอน ก็ให้ใช้สูตรคำนวณหาจำนวนตัวอย่าง เช่น สูตรของ Taro Yamane และ Krejcie and Morgan สามารถดูได้จากผลการคำนวณ (เชื่อมั่น 95%) ที่มีปรากฏตามตำราบางเรื่องที่ใช้วิธีการคำนวณจากประชากรเป็นร้อยเป็นพันเป็นหมื่นและเป็นแสนหรือล้านให้เป็นตัวอย่างลงไว้ในตาราง ทั้งนี้ เพื่อความสะดวกต่อผู้วิจัยและสามารถนำประชากรมาเปิดเทียบหาจำนวนตัวอย่างเพื่อนำไปใช้ได้ทันที

3. การศึกษาต้นทุนและผลตอบแทนจากการลงทุน การใช้ตัวอย่างหรือกลุ่มตัวอย่างไม่จำเป็นจะต้องใช้ตามขนาดตัวอย่าง (Sample Size) ตามสูตรของนักสถิติที่นิยมใช้กันในปัจจุบันหรือการสุ่มตัวอย่าง (Random Sampling) โดยสุ่มตัวอย่างเพียง 3 - 5 ราย ทั้งนี้เพราะลักษณะของต้นทุนคงที่และต้นทุนผันแปรจะไม่ค่อยแตกต่างกัน เพียงแต่การจำแนกกิจการออกเป็นขนาดเล็กขนาดกลางและใหญ่ เพื่อวิเคราะห์ผลตอบแทนอันเนื่องมาจากขนาดของปริมาณการผลิต

ตารางที่ 5.1 แสดงประชากรและขนาดตัวอย่างที่คำนวณจากสูตรของ Taro Yamane (1973) เชื่อมั่น 95%

ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง	ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง	ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง	ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง
20	19	180	124	600	240	8000	380
25	23	190	129	650	248	9000	383
30	28	200	134	700	255	10000	385
35	32	210	138	750	261	12000	387
40	36	220	142	800	267	15000	390
45	41	230	146	850	272	20000	392
50	45	240	150	900	277	25000	394
55	48	250	154	950	282	30000	395
60	52	260	158	1000	286	40000	396
65	56	270	161	1100	294	50000	397
70	60	280	165	1300	306	60000	397
75	63	290	168	1500	316	80000	398
80	67	300	172	1700	324	100000	398
85	70	320	178	1900	331	120000	398
90	74	340	184	2100	336	150000	398
95	77	360	190	2400	343	200000	399
100	80	380	195	2700	348	300000…	399
110	86	400	200	3000	353	400000	400
120	94	420	205	3500	359	500000	400
130	98	440	210	4000	364	700000	400
140	104	460	214	4500	367	900000	400
150	109	480	218	5000	370	1000000	400
160	114	500	222	6000	375	1500000	400
170	120	550	231	7000	378	2000000	400

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

ที่มา: Yamane Taro (1973)

ตารางที่ 5.2 แสดงประชากรและขนาดตัวอย่างที่คำนวณจากสูตรของอภินันท์ จันตะนี (2549) เชื่อมั่น 93%

ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง	ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง	ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง	ประชากรทั้งหมด	จำนวนกลุ่มตัวอย่าง
50	22	380	98	900	140	10,000	196
60	26	400	101	940	142	12,000	197
70	30	420	104	960	143	14,000	198
80	34	440	106	980	144	16,000	199
90	37	460	108	1,000	145	18,000	200
100	40	480	110	1,100	149	20,000	200
110	43	500	112	1,200	152	24,000	201
120	46	520	114	1,300	155	26,000	201
130	49	540	116	1,400	158	28.,000	201
140	52	560	118	**1,500**	**160**	30,000	201
150	54	580	120	1,800	166	35,000	202
160	58	600	122	2,000	170	40,000	202
170	60	620	123	2,500	176	45,000	202
180	62	640	125	3,000	180	50,000	202
190	64	660	126	3,500	183	55,000	202
200	67	680	127	4,000	185	**60,000**	**203**
220	71	700	129	4,500	187	65,000	203
240	75	720	130	5,000	189	70,000	203
260	79	740	132	5,500	190	75,000	203
280	83	780	134	6,000	191	80,000	203
300	86	800	135	6,500	192	85,000	203
320	89	840	137	7,000	193	90,000	203
340	92	860	138	8,000	194	95,000	203
360	95	880	139	9,000	195	100,000	203

ที่มา: การพัฒนาสูตรเพื่อให้นักศึกษาได้ฝึกปฏิบัติวิจัยที่ใช้ตัวอย่างไม่มากนัก พ.ศ. 2549 $n = \frac{N}{2 + N(d)^2}$

สรุป

สถิติที่นำมาใช้ในกระบวนการทำวิจัย นับว่ามีความสำคัญมาก เนื่องจากสถิติจะช่วยในกระบวนการวิจัยหลายประการอาทิ การทดสอบเครื่องมือเพื่อการวิจัย การสุ่มตัวอย่าง การแจกแจงประชากร การสำรวจข้อมูล และการวิเคราะห์ข้อมูล รวมทั้งการประเมินค่าต่าง ๆ ให้ออกมาเป็นตัวเลข นอกจากนั้นสถิติ ยังเป็นประโยชน์สำหรับการวิเคราะห์เหตุการณ์ต่าง ๆ หรือวิเคราะห์ข้อมูลเพื่อทดสอบสมมติฐานในการวิจัยทางการบริหารซึ่งหากการวิจัยผ่านกระบวนการที่ถูกต้องและเลือกใช้สถิติที่เหมาะสมแล้วจะช่วยสร้างมั่นใจให้กับผู้ใช้งานวิจัยต่อไป