บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงเนื้อหาความรู้และงานวิจัยที่เกี่ยวข้อง โดยผู้วิจัยได้แบ่งออกเป็น 6 ตอน ดังนี้

1. การสุ่มตัวอย่าง (Sampling)
2. การสุ่มตัวอย่างแบบแบ่งชั้น (Stratified Sampling)
3. ความคลาดเคลื่อนที่เกิดจากการสำรวจ (Component of surver error)
4. การประมาณค่า (Estimation)
5. การแจกแจงค่าสถิติที่ได้จากการสุ่มตัวอย่าง (Sampling Distribution)
6. งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 การสุ่มตัวอย่าง

ในการสำรวจเพื่อหาข้อมูลทางสถิติระดับใหญ่ซึ่งเรียกกันว่า Large Scale Statistical Survey เช่น การสำรวจผลผลิตการเกษตรของประเทศ การสำรวจแรงงานของประชากร การสำรวจเพื่อหาข้อมูลทางเศรษฐกิจและสังคม ฯลฯ อาจแบ่งระเบียบวิธีการสำรวจ (Survey methodology) ออกเป็น 2 วิธีใหญ่ๆ (นิยม ปุราคำ, 2517) คือ

1.การสำรวจโดยวิธีแจงนับอย่างครบถ้วน หรือ การสำมะโน (Complete enumeration method or Census) เช่น การเก็บรวบรวมข้อมูลเกี่ยวกับจำนวนและคุณลักษณะของประชากรและเคหะสถาน ซึ่งเรียกกันว่า สำมะโนประชากรและเคหะ ซึ่งรัฐบาลจัดทำโดยจัดส่งพนักงานแจงนับออกไปยังครัวเรือนทุกครัวเรือน เพื่อสอบถามข้อมูลต่างๆเกี่ยวกับประชากรทุกคนและบันทึกลงในแบบสำมะโนที่กำหนดขึ้น การเก็บรวบรวมข้อมูลสถิติโดยวิธีนี้ ต้องลงทุนคือใช้กำลังคน และงบประมาณมาก ซึ่งอาจทำให้เกิดปัญหาในการหาปัจจัยเหล่านี้ให้พอเพียงแก่ความต้องการของการสำรวจนั้น ถึงแม้ว่าเราจะสามารถหางบประมาณที่จะต้องใช้มาได้ทั้งหมด แต่ก็ยังมีปัญหาในด้านเวลาเนราะการสำรวจโดยวิธีนี้มักจะต้องใช้เวลานานทำให้ข้อมูลที่ได้ขาดความทันสมัย ไม่ทันต่อความต้องการของผู้ที่

018853

ต้องการข้อมูลในการศึกษาเรื่องที่เร่งด่วน ถ้าจะกล่าวโดยสรุปแล้วจะเห็นได้ว่าการสำรวจสถิติโดยการสำมะโนนั้นจะมีข้อเสียอยู่หลายอย่าง คือ ใช้กำลังคนมาก ใช้งบประมาณมาก ใช้เวลามากยากในการควบคุมงานซึ่งจะต้องใช้คนสำรวจและเก็บข้อมูลจำนวนมากซึ่งเป็นผลทำให้ข้อมูลที่ได้มีโอกาสผิดพลาดได้มาก

2.การสำรวจโดยวิธีการแจงนับเพียงบางส่วน หรือ การสุ่มสำรวจ (Partial enumeration method or Sampling survey) ระเบียบวิธีการสำรวจนี้ถูกพัฒนาขึ้นเพื่อแก้ปัญหาและข้อเสียของการสำมะโนที่กล่าวมาแล้ว ทั้งนี้โดยอาศัยทฤษฎีทางคณิตศาสตร์สถิติ (Mathematical statistics) เกี่ยวกับความน่าจะเป็น (Probability) การแจกแจงตัวแปรสุ่ม (Distribution of random variables) ประกอบกับการวางแบบการสำรวจ (Survey design) และการประมาณผล (Estimation) วิธีนี้เป็นการเลือกตัวแทนของประชากรที่เราต้องการศึกษาขึ้นมาจำนวนหนึ่งแล้วใช้ข้อมูลที่ได้จากตัวอย่าง ประมาณตัวเลขที่ต้องการโดยอาศัยหลักการประมาณผลทางสถิติที่เหมาะสม จะทำให้เราได้ค่าประมาณของตัวเลขที่ใกล้เคียงกับค่าที่ได้จากประชากรทั้งหมด ข้อได้เปรียบของระเบียบวิธีการสำรวจโดยการสุ่มสำรวจทั้งในทางทฤษฎีและทางปฏิบัติก็คือ ใช้กำลังคนน้อย ใช้งบประมาณในการสำรวจน้อย ใช้เวลาในการดำเนินการทั้งในด้านงานสนามและการประมวลผลน้อย สามารถควบคุมคุณภาพของข้อมูลได้ สามารถศึกษาและวิเคราะห์ปัญหาได้ลึกและสมบูรณ์กว่าการศึกษาข้อมูลที่มีจำนวนมาก

ในปัจจุบันระเบียบวิธีการสำรวจจากตัวอย่างจัดว่าเป็นเครื่องมือที่สำคัญที่สุดในการหาข้อมูลสถิติต่างๆของรัฐบาลและเอกชน ไม่ว่าจะเป็นการสำรวจข้อมูลทางด้านการเกษตร อุตสาหกรรม สาธารณสุข การคมนาคม การศึกษา และข้อมูลทางเศรษฐกิจ สังคม อื่นๆ รวมทั้งการหยั่งเสียงประชามติ การวิจัยตลาด อย่างไรก็ตาม มีข้อควรระวังในการศึกษาข้อมูลที่ได้จากการสุ่มตัวอย่างที่ควรจะต้องคำนึงถึงมีดังนี้

2.1 ถ้าการสุ่มตัวอย่างนั้นไม่ได้รับการออกแบบวิธีการสุ่มตัวอย่าง และนำไปปฏิบัติอย่างถูกต้องแล้วผลที่ได้จะมีความคลาดเคลื่อน และนำไปสู่การสรุปที่ผิดพลาดได้

2.2 การสุ่มตัวอย่างแต่ละวิธีนั้นต้องอาศัยเทคนิคเฉพาะของแต่ละวิธี และมีข้อจำกัดในแต่ละวิธีซึ่งจะต้องคำนึงถึงความเหมาะสมกับเนื้อเรื่องที่ต้องทำการศึกษานั้นด้วย

2.3 การเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างประชากรจะต้องมีการตระเตรียมรายละเอียดเกี่ยวกับประชากร เช่น บัญชีรายชื่อ การจำแนกแยกประเภทประชากร หน่วยการสุ่ม

วิธีการสุ่ม เป็นต้น

2.4 ถ้าใช้ขนาดตัวอย่างไม่มากพอ จะมีปัญหาในแง่ของการสรุปอ้างอิงไปยังประชากรตลอดจนไม่สามารถประมาณค่าตัวเลขในระดับย่อยๆได้

มโนทัศน์พื้นฐานที่สำคัญ

มโนทัศน์พื้นฐานสำคัญที่เกี่ยวกับการสุ่มตัวอย่าง มีดังนี้ (ศิริชัย กาญจนวาสี, 2533)

1. ประชากร (Population) หมายถึง กลุ่มของสิ่งต่างๆทั้งหมดที่สนใจศึกษา ซึ่งอาจจะเป็นคน สิ่งของ หรือเหตุการณ์ต่างๆ ค่าที่คำนวณได้จากข้อมูลทุกหน่วยของประชากร เรียกว่า ค่าพารามิเตอร์ (Parameter) จึงเป็นค่าที่เป็นจริง และถือว่าเป็นค่าที่ถูกต้อง เช่น ค่าเฉลี่ยเลขคณิต (μ) ส่วนเบี่ยงเบนมาตรฐาน (σ) สัดส่วน (Π) สหสัมพันธ์ (ρ) เป็นต้น

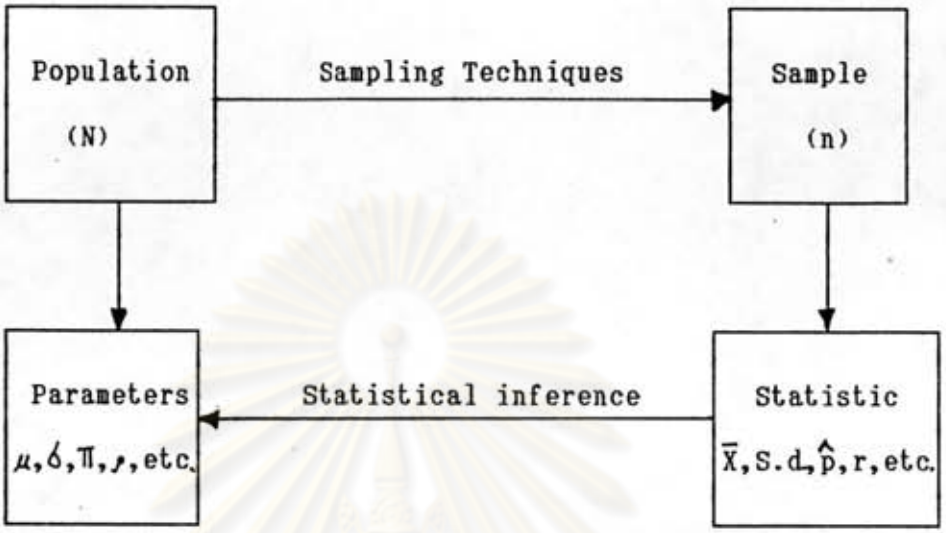
2. กลุ่มตัวอย่าง (Sample) หมายถึง ส่วนหนึ่งของประชากรที่ถูกเลือกหรือสุ่มเพื่อใช้ศึกษาแทนประชากร ค่าที่คำนวณได้จากข้อมูลกลุ่มตัวอย่างเรียกว่า ค่าสถิติ (Statistics) จึงเป็นค่าที่เป็นจริงเฉพาะกลุ่มตัวอย่างนั้น โดยทั่วไปเป็นค่าที่มีความคลาดเคลื่อนจากค่าพารามิเตอร์

3. การสุ่มตัวอย่าง (Sampling Techniques) หมายถึง การกระทำเพื่อให้ได้มาซึ่งกลุ่มตัวอย่างเพื่อใช้ศึกษาแทนประชากร การออกแบบการสุ่มตัวอย่างที่ดีประกอบด้วย วิธีการสุ่มตัวอย่างที่ดี คือไม่ลำเอียง ปราศจากอคติ และขนาดของกลุ่มตัวอย่างที่เหมาะสม (ขนาดกลุ่มตัวอย่างที่ถูกต้องตามหลักสถิติ คือ มีความเป็นตัวแทนประชากร ตลอดจนเป็นขนาดกลุ่มตัวอย่างที่พอเหมาะตามหลักปฏิบัติ คือ มีความเป็นไปได้ในการติดตามเก็บรวบรวมข้อมูล)

4. การสรุปอ้างอิงทางสถิติ (Statistical Inference) หมายถึง การวิเคราะห์ทางสถิติ เพื่อใช้ค่าสถิติที่คำนวณได้จากกลุ่มตัวอย่างไปทำการประมาณค่า (Estimation) หรือทดสอบสมมติฐาน (Hypothesis testing) เกี่ยวกับค่าพารามิเตอร์ของประชากร

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพที่ 1 มโนทัศน์ที่สำคัญเกี่ยวกับการสุ่มตัวอย่างและการวิเคราะห์ทางสถิติ



ขั้นตอนในการสุ่มกลุ่มตัวอย่าง

ในเชิงปฏิบัติแล้วผู้วิจัยอาจวางแผนในการที่จะเลือกสุ่มกลุ่มตัวอย่างที่ดีได้ตามลำดับขั้นดังนี้

1.พิจารณาความมุ่งหมายของการวิจัย หรือการเลือกกลุ่มตัวอย่าง ผู้วิจัยจะต้องวิเคราะห์จุดมุ่งหมายของการวิจัยอย่างละเอียดและให้ชัดเจนในจุดมุ่งหมาย หรือปัญหาของการวิจัยเพื่อเป็นการหาลักษณะของกลุ่มตัวอย่างหรือประชากรที่ถูกต้องที่สุดว่าจะศึกษากับใคร มีคุณสมบัติเช่นไร

2.ให้คำจำกัดความของประชากรที่จะใช้เพื่อสนองจุดมุ่งหมายของการวิจัยนั้น ผู้วิจัยจะต้องให้ความหมายของประชากรได้ว่าหมายถึงใคร ครอบคลุมกว้างขวางหรือมีขอบเขตแค่ไหน มีคุณลักษณะที่จำกัดอย่างไร เช่น อาจหมายถึงนักเรียนที่เรียนชั้น ม.3 ทั่วประเทศ หรือเฉพาะนักเรียนที่เรียนในโรงเรียนสาธิตของมหาวิทยาลัยหรือวิทยาลัยครูเท่านั้น เป็นต้น

3.การกำหนดหน่วยของตัวอย่าง (Sampling unit) ก่อนที่จะเลือกกลุ่มตัวอย่างผู้วิจัยจะต้องกำหนดหน่วยของตัวอย่างก่อน โดยพิจารณาจากคุณลักษณะของประชากรที่เราจะศึกษา ส่วนใหญ่แล้วหน่วยของตัวอย่างนี้มักจะหมายถึง หน่วยหรือส่วนที่ย่อยที่สุดของประชากรในการเลือกแต่ละครั้ง ซึ่งการเลือกหน่วยย่อยที่สุดนี้ บางครั้งก็ทำได้ง่ายบางครั้งก็ทำได้ยาก ทั้งนี้ขึ้นอยู่กับหน่วยของตัวอย่างนั้นว่าชัดเจนและสะดวกในการเลือกแค่ไหน ฉะนั้น ก่อนเลือกกลุ่มตัวอย่างผู้วิจัยจะต้องกำหนดหน่วยของตัวอย่างให้ได้เสียก่อน

4.กำหนดขอบข่ายของประชากร การกำหนดขอบข่ายของประชากร ก็คือ การรวบรวมรายชื่อ หรือทำบัญชีหน่วยของตัวอย่าง (Sampling unit) หรือเป็นการหาขนาดของประชากรนั่นเอง

5.ประมาณขนาดของตัวอย่าง (Estimated sample size) ดังที่กล่าวมาแล้วว่าการเลือกกลุ่มตัวอย่างนั้นต้องคำนึงถึงว่ากลุ่มตัวอย่างที่ได้นั้นจะต้องเป็นตัวแทนที่ดีของประชากร ฉะนั้น การเลือกกลุ่มตัวอย่างมาศึกษานั้นจะต้องให้มีขนาด หรือจำนวน ที่เป็นสัดส่วนพอเหมาะกับขนาดของประชากร

6.เลือกวิธีการสุ่มกลุ่มตัวอย่างที่เหมาะสม (Sampling technique) การเลือกกลุ่มตัวอย่างที่ดีและเหมาะสมนั้นขึ้นอยู่กับผู้วิจัยมีความรู้เกี่ยวกับประชากรนั้นมากน้อยเพียงใด ซึ่งถ้าเรารู้เรื่องประชากรที่เราจะวิจัยมาก โอกาสที่เราจะเลือกกลุ่มตัวอย่างได้ตรงกับวัตถุประสงค์ของการวิจัยก็มีมากขึ้น

<u>วิธีการสุ่มตัวอย่าง</u> จำแนกออกเป็น 2 วิธีใหญ่ (ศิริชัย กาญจนวาสี, 2526) ดังนี้

1. การสุ่มที่ไม่คำนึงถึงความน่าจะเป็นในการสุ่ม (Non-probability Sampling) การสุ่มแบบนี้ไม่คำนึงถึงว่ากลุ่มตัวอย่างที่ได้รับเลือกมานั้น จะมีความน่าจะเป็นหรือโอกาสที่จะได้รับเลือกนั้นเป็นเท่าใด เป็นการสุ่มตัวอย่างที่ขึ้นอยู่กับการควบคุมหรือการตัดสินของผู้วิจัยเป็นอย่างมาก การสุ่มแบบนี้ไม่สามารถประกันได้ว่าสมาชิกทุกหน่วยจากกลุ่มประชากรนั้นจะมีโอกาสได้รับเลือกมาเป็นสมาชิกในกลุ่มตัวอย่าง การสุ่มแบบนี้จะทำให้เกิดความลำเอียงในการสุ่มตัวอย่างได้อย่างง่ายดาย วิธีการสุ่มตัวอย่างที่ไม่คำนึงถึงความน่าจะเป็นในการสุ่ม ได้แก่

1.1 การสุ่มโดยการบังเอิญ (Accidental Sampling)

1.2 การสุ่มแบบโควตา (Quota Sampling)

1.3 การสุ่มอย่างเฉพาะเจาะจง (Purposive Sampling)

1.4 การสุ่มแบบลูกโซ่ (Snowball Sampling)

1.5 การสุ่มตัวอย่างแบบเฉพาะที่มีข้อมูลอยู่แล้ว (Haphazard Sampling)

การสุ่มตัวอย่างที่ไม่คำนึงถึงความน่าจะเป็นในการสุ่มนั้น สะดวกในกรณีที่กลุ่มตัวอย่างที่จะเลือกมานั้นมีขนาดเล็กและผู้วิจัยต้องการแนวคิดเกี่ยวกับลักษณะต่างๆ ของกลุ่มประชากรในระยะเวลาอันจำกัด การสุ่มตัวอย่างในลักษณะนี้มีข้อจำกัดดังต่อไปนี้

1) ผลการวิจัยที่ได้ไม่สามารถสรุปอ้างอิงไปสู่ประชากรทั้งหมดได้ จะสรุปอยู่ในขอบเขตของกลุ่มตัวอย่างเท่านั้น ข้อสรุปนั้นจะสรุปกลับไปหากลุ่มประชากรได้ก็ต่อเมื่อ กลุ่มตัวอย่างมีลักษณะต่างๆ ที่สำคัญๆ เหมือนกับลักษณะของกลุ่มประชากร

2) กลุ่มตัวอย่างที่ได้นั้นขึ้นอยู่กับการตัดสินของผู้วิจัยและองค์ประกอบบางตัวที่ไม่สามารถควบคุมได้ ซึ่งทำให้ไม่สามารถที่จะคำนวณหาค่าความคลาดเคลื่อนที่เกิดจากการสุ่ม (Sampling error) โดยการสุ่มวิธีนี้ได้

2. การสุ่มที่คำนึงถึงความน่าจะเป็นในการสุ่ม (Probability sampling) เป็นการสุ่มตัวอย่างที่คำนึงถึงความน่าจะเป็น หรือโอกาสของสมาชิกแต่ละหน่วยที่จะได้รับเลือก ซึ่งสมาชิกทุกๆ หน่วยของกลุ่มประชากรจะมีความน่าจะเป็นหรือโอกาสที่จะได้รับเลือกเท่าเทียมกัน กลุ่มตัวอย่างที่ได้รับการสุ่มแบบนี้จะเป็นตัวแทนที่ดีของกลุ่มประชากรเป้าหมายได้ดีกว่ากลุ่มตัวอย่างที่สุ่มแบบไม่คำนึงถึงความน่าจะเป็นในการสุ่ม วิธีการสุ่มตัวอย่างที่คำนึงถึงความน่าจะเป็นในการสุ่มได้แก่

2.1 การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) เป็นวิธีการสุ่มตัวอย่างที่เปิดโอกาสให้แต่ละหน่วยย่อยที่สุดของประชากรมีสิทธิได้รับการเลือกเป็นสมาชิกของกลุ่มตัวอย่างเท่าเทียมกันและเป็นอิสระจากกัน

การสุ่มตัวอย่างแบบง่ายสามารถกระทำได้โดยวิธีการต่างๆ ดังนี้

1) วิธีจับฉลาก

2) วิธีใช้ตารางเลขสุ่ม

3) วิธีใช้คอมพิวเตอร์

การสุ่มตัวอย่างแบบง่ายนี้เหมาะสำหรับใช้กับประชากรที่มีลักษณะที่สนใจศึกษาคล้ายคลึงกัน วิธีนี้มีความสะดวกในการสุ่ม การคำนวณขนาดกลุ่มตัวอย่างและสูตรสถิติต่างที่เกี่ยวข้องไม่ยุ่งยาก แต่อาจจะสิ้นเปลืองทรัพยากรในแง่ค่าใช้จ่ายและเวลาในการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างที่อยู่กันอย่างกระจาย

2.2 การสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) เป็นวิธีการสุ่มตัวอย่างจากบัญชีรายชื่อสมาชิกของประชากร โดยทำการสุ่มเป็นช่วงๆ อย่างเป็นระบบ วิธีการสุ่มแบบนี้จึงน่าไปปฏิบัติได้สะดวกกว่าการสุ่มตัวอย่างแบบง่าย เมื่อประชากรมีขนาดใหญ่และมีบัญชีรายชื่อสมาชิกของประชากร นอกจากนี้ ควรใช้กับประชากรที่มีสมาชิกจัดเรียงกันอย่างสุ่ม

2.3 การสุ่มตัวอย่างแบบแบ่งชั้น (Stratified Sampling) เป็นวิธีการสุ่ม

ตัวอย่างที่มีการจำแนกประชากรออกเป็นพวกหรือชั้น (Stratum) โดยให้ประชากรในชั้นเดียวกันมีความคล้ายคลึงกัน(Homogeneous) แต่ประชากรที่ต่างชั้นกันมีความแตกต่างกัน(Heterogeneous) การสุ่มตัวอย่างแบบแบ่งชั้นนี้มีความยืดหยุ่นและเอื้อประโยชน์ได้สูงในทางปฏิบัติเพราะผู้วิจัยสามารถเลือกวิธีการสุ่มและกำหนดขนาดกลุ่มตัวอย่างสำหรับแต่ละชั้น(stratum) ให้มีความแตกต่างกันได้ตามความเหมาะสมของลักษณะประชากรย่อยของแต่ละชั้น และยังช่วยให้ผู้วิจัยสามารถทำการวิเคราะห์และสรุปผลการเปรียบเทียบระหว่างประชากรกลุ่มย่อยได้อย่างมั่นใจ นอกจากนี้ยังมีผลดีทางทฤษฎีในแง่ของการช่วยลดความคลาดเคลื่อนของการประมาณค่าและการทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์ได้ต่ำลงกว่าวิธีการสุ่มตัวอย่างแบบง่ายและแบบมีระบบ

2.4 การสุ่มตัวอย่างแบบตามกลุ่ม (Cluster Sampling) เป็นวิธีการสุ่มตัวอย่างที่นิยมใช้กับประชากรที่อยู่รวมกันเป็นกลุ่ม (cluster)โดยที่แต่ละกลุ่มประชากรประกอบด้วยสมาชิกภายในกลุ่มที่มีลักษณะหลากหลาย แต่เมื่อพิจารณาระหว่างกลุ่มประชากรจะประกอบด้วยสมาชิกโดยส่วนรวมมีลักษณะคล้ายคลึงกัน (Homogeneous) การสุ่มตัวอย่างแบบตามกลุ่มนี้มีความความสะดวกในกรณีที่ไม่จำเป็นต้องมีกรอบการสุ่มของสมาชิกที่สมบูรณ์ และลดค่าใช้จ่ายในการเก็บรวบรวมข้อมูล วิธีการสุ่มแบบนี้มีความเหมาะสมเมื่อกลุ่มประชากรแต่ละกลุ่มมีความคล้ายคลึงกัน และมีลักษณะต่างๆของประชากรครบถ้วนแต่มีจุดอ่อนในแง่ความยากของการจัดหรือหากลุ่มตามธรรมชาติที่มีความหลากหลายภายในกลุ่ม แต่ระหว่างกลุ่มประชากรมีความคล้ายคลึงกัน และมีความยุ่งยากในการคำนวณขนาดกลุ่มตัวอย่างและค่าสถิติต่างๆที่เกี่ยวข้อง นอกจากนี้ยังอาจเกิดความคลาดเคลื่อนของการประมาณค่าหรือทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์ได้สูงกว่าวิธีการสุ่มแบบง่าย

2.5 การสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage Sampling) การสุ่มตัวอย่างแบบหลายขั้นตอนเป็นวิธีการสุ่มตัวอย่างที่ประกอบด้วยการสุ่มตั้งแต่ 2 ขั้นขึ้นไป วิธีนี้นิยมใช้กับประชากรที่มีโครงสร้างแบ่งเป็นระดับชั้นต่างๆกัน เช่น ภาค จังหวัด อำเภอ ตำบล หมู่บ้าน ครัวเรือน หรือ เขตการศึกษา กลุ่มโรงเรียน โรงเรียน ห้องเรียน เป็นต้น วิธีการสุ่มเริ่มต้นด้วยการสุ่มในระดับชั้นใหญ่ขึ้นมาก่อน จากชั้นใหญ่ที่สุ่มได้ ทำการสุ่มชั้นย่อยที่อยู่ระดับรองลงไป จนถึงหน่วยที่ต้องการศึกษาโดยให้มีจำนวนสมาชิกรวมเท่ากับขนาดกลุ่มตัวอย่างที่ต้องการ การสุ่มแบบหลายขั้นตอนนี้เป็นเทคนิคที่ผสมวิธีสุ่มหลายวิธีเข้าด้วยกัน ข้อดีการสุ่มแบบหลายขั้นตอน คือ สามารถใช้กับประชากรที่มีจำนวน หรือขอบเขตที่กว้างขวางจนไม่สามารถหาบัญชีรายชื่อสมาชิก หรือ กรอบการสุ่มที่สมบูรณ์ได้ จึงทำการสุ่มตามระดับชั้นจนถึงกลุ่มตัวอย่างที่ต้องการได้ นอกจากนี้ ยังเป็นวิธีการสุ่ม

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

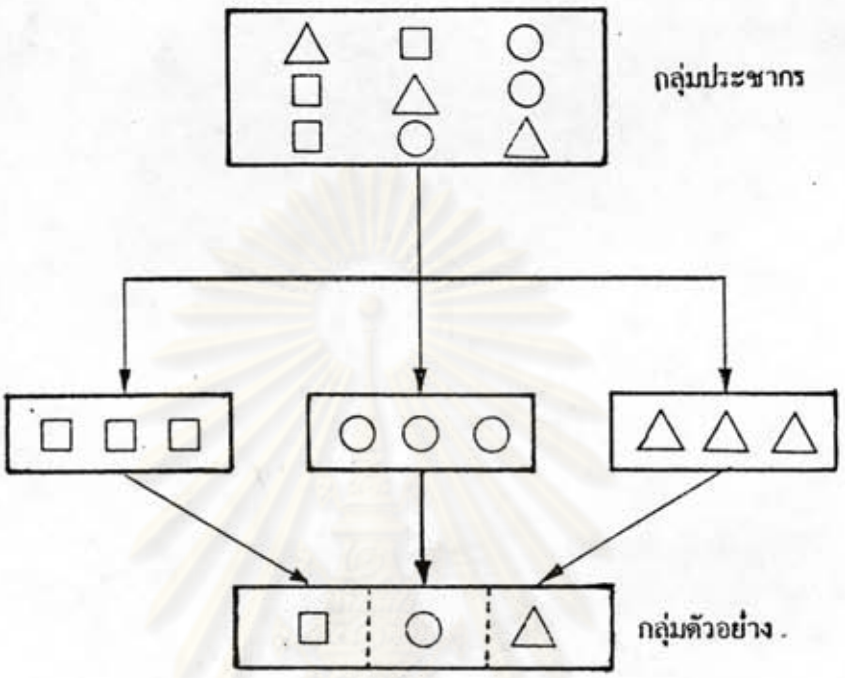
ที่สะดวกและง่ายต่อการนำไปปฏิบัติ แต่มีจุดอ่อนคล้ายๆวิธีการสุ่มแบบตามกลุ่ม กล่าวคือ มีความยุ่งยากในการคำนวณขนาดกลุ่มตัวอย่างของชั้นต่างๆ สูตรสถิติที่เกี่ยวข้องค้านข้างสลับซับซ้อน และอาจต้องใช้ขนาดกลุ่มตัวอย่างมากกว่าวิธีการสุ่มวิธีอื่นๆจึงจะให้ประสิทธิภาพพอๆกัน รวมทั้งกลุ่มตัวอย่างที่ได้จะอยู่กันอย่างกระจัดกระจาย

การวางแผนการสุ่มตัวอย่างประชากรแต่ละวิธีมีข้อดีและข้อเสียอยู่ในตัวของมันเอง ซึ่งจะต้องนิจารณาให้รอบคอบ และเลือกวิธีที่ดีที่สุด เหมาะสมที่สุด เพื่อให้ได้ตัวแทนของประชากรที่สอดคล้อง เหมาะสมกับสภาพที่ทำการศึกษา อย่างไรก็ตามได้มีผู้ทำการศึกษาวิจัยเพื่อเปรียบเทียบประสิทธิภาพและความแม่นยำในการประมาณค่าพารามิเตอร์ของวิธีการสุ่มทั้ง 5 วิธีโดยใช้เทคนิคการชิมมูเลชัน (Simmulation) พบว่า การสุ่มแบบแบ่งชั้นเป็นวิธีที่มีประสิทธิภาพมากที่สุด ซึ่งจากสาเหตุดังกล่าวจึงทำให้ผู้วิจัยเลือกศึกษาการสุ่มแบบแบ่งชั้นในการวิจัยนี้

ตอนที่ 2 การสุ่มตัวอย่างแบบแบ่งชั้น (Stratified Sampling)

ในงานสำรวจข้อมูลนั้นผู้วิจัยมักจะพบกับปัญหาในทางปฏิบัติเกี่ยวกับการกระจัดกระจายของหน่วยสำรวจทำให้การควบคุมงานในสนามเป็นไปได้โดยยาก บางครั้งเมื่อใช้การสุ่มตัวอย่างแบบง่าย ปรากฏว่ากลุ่มตัวอย่างไม่ใช่ตัวแทนที่ดีพอ เนื่องจากมีบางส่วนของประชากรที่ไม่ได้รับเลือก หรือได้รับเลือกเข้ามาในปริมาณที่น้อยมากทำให้ค่าประมาณที่ได้มีค่าต่ำกว่าหรือสูงกว่าความเป็นจริง และในบางครั้งผู้วิจัยต้องการทราบข้อมูลจากประชากรย่อยบางกลุ่ม แต่ขาดข้อมูลที่เพียงพอเนื่องจากกลุ่มประชากรย่อยที่สนใจนั้นได้รับการเลือกมาเป็นตัวอย่างน้อยเกินไปหรือไม่ปรากฏเลย ด้วยปัญหาดังกล่าวจึงได้มีการพัฒนาแบบแผนการสุ่มตัวอย่างขึ้นมาใหม่ เรียกว่า "แผนการสุ่มตัวอย่างแบบแบ่งชั้น" แผนดังกล่าวมีขั้นตอนในการสุ่มด้วยการจำแนกประชากรออกเป็นกลุ่มประชากรย่อยที่ไม่ซ้ำซ้อนกัน(Nonoverlapping Subpopulation) เรียกว่า ชั้นภูมิ (Strata) โดยพยายามจัดให้หน่วยตัวอย่างที่มีลักษณะคล้ายคลึงกันหรือเนื้อเดียวกัน (Homogeneous)ไว้ในชั้นเดียวกัน และหน่วยตัวอย่างที่มีความแตกต่างกันจัดให้อยู่ต่างชั้นภูมิกัน จากนั้นจึงดำเนินการสุ่มหน่วยตัวอย่างมาจากทุกชั้นภูมิในปริมาณมากน้อยแตกต่างกันไปตามความเหมาะสม ซึ่งในการสุ่มตัวอย่างในแต่ละชั้นภูมิเป็นไปโดยอิสระจากกัน ลักษณะการจัดชั้นภูมิอาจจะแสดงโดยใช้แผนภาพประกอบได้ดังนี้ (บุญเรือง ขจรศิลป์, 2530)

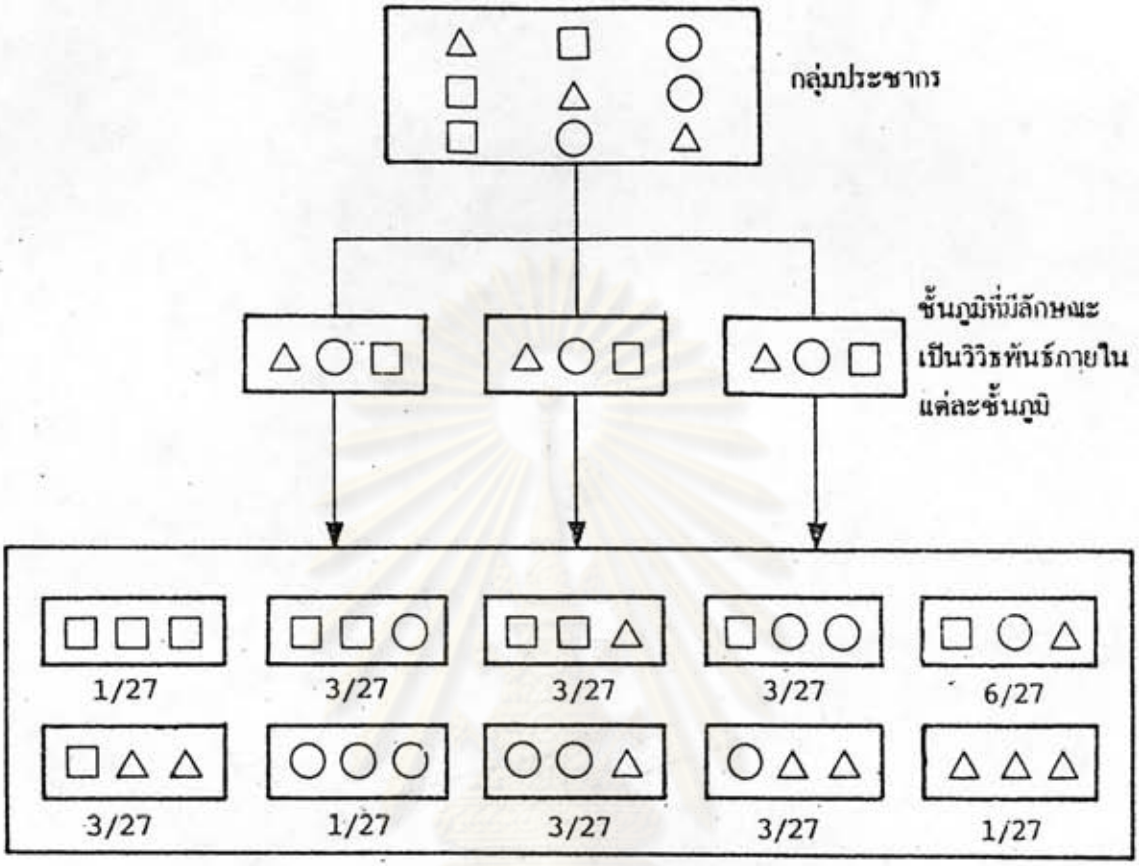
แผนภาพที่ 2 วิธีการสุ่มตัวอย่างแบบแบ่งชั้น



จากแผนภาพแสดงให้เห็นว่าประชากรประกอบด้วยสมาชิกที่มีลักษณะแตกต่างกัน 3 ลักษณะ คือ □ ○ △ ในการจัดชั้นภูมิจะต้องพยายามจัดให้ภายในแต่ละชั้นภูมิมีลักษณะเหมือนกันมากที่สุด และมีความแตกต่างระหว่างชั้นภูมิให้มากที่สุด ดังนั้นจึงจัดแยกเป็น 3 ชั้นภูมิ คือ ชั้นภูมิที่มีสมาชิก □ ชั้นภูมิที่มีสมาชิก ○ ชั้นภูมิที่มีสมาชิก △ หลังจากนั้นผู้วิจัยจึงสุ่มสมาชิกจากแต่ละชั้นภูมิ ในที่นี้สมมติให้สุ่มมาชั้นละ 1/3 ของแต่ละชั้นภูมิซึ่งจะได้กลุ่มตัวอย่างซึ่งประกอบด้วยสมาชิกที่มีลักษณะ □ ○ และ △ เหมือนกับลักษณะของประชากร

ข้อผิดพลาดที่อาจจะเกิดขึ้นได้ คือ ผู้วิจัยจัดให้ภายในแต่ละชั้นภูมิมีลักษณะที่แตกต่างกันและระหว่างชั้นภูมิมีลักษณะที่เหมือนกัน ผลที่ตามมาก็คือผู้วิจัยจะได้กลุ่มตัวอย่างที่มีลักษณะแตกต่างจากลักษณะของกลุ่มประชากร ดังแสดงในแผนภาพต่อไปนี้

แผนภาพที่ 3 ลักษณะการจัดชั้นภูมิที่ผิดพลาดในการสุ่มตัวอย่างแบบแบ่งชั้น



กลุ่มตัวอย่างที่เป็นไปได้

จากแผนภาพแสดงให้เห็นถึงความผิดพลาดที่เกิดขึ้นในการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ ผลที่ได้เราจะพบว่าโอกาสของกลุ่มตัวอย่างที่จะเกิดจากการสุ่มทั้งหมดมี 27 กลุ่มตัวอย่าง ซึ่งจะมีลักษณะที่แตกต่างกัน 10 ลักษณะ แต่จากแผนภาพจะเห็นได้ว่าจากกลุ่มตัวอย่างทั้งหมด 27 กลุ่มตัวอย่างนั้น มีโอกาสที่จะได้กลุ่มตัวอย่างที่มีลักษณะเหมือนกับกลุ่มประชากรเพียง 6 กลุ่มเท่านั้น (6/27) นอกนั้นอีก 21 กลุ่ม (21/27) จะมีลักษณะไม่เหมือนกับกลุ่มประชากร ซึ่งโอกาสผู้วิจัยจะได้กลุ่มตัวอย่างที่มีลักษณะต่างจากกลุ่มประชากรมีมากกว่า

<u>การประมาณค่ามัชฌิมเลขคณิตและค่าความแปรปรวน</u>

1. <u>ค่ามัชฌิมเลขคณิต</u> เราสามารถประมาณค่าได้จากสูตร (Cochran, 1977)

$$\bar{X}_{st} = \frac{\sum_{h=1}^{L} N_h \cdot \bar{X}_h}{N}$$

โดยที่

$$\bar{X}_h = \frac{\sum_{i=1}^{n_h} X_{hi}}{n_h}$$

เมื่อ $\bar{X}_{st}$ คือ ค่ามัชฌิมเลขคณิตของกลุ่มตัวอย่าง

$\bar{X}_h$ คือ ค่ามัชฌิมเลขคณิตของกลุ่มตัวอย่างในชั้นภูมิที่ h

X_{hi} คือ คะแนนของสมาชิกลำดับที่ i ในชั้นภูมิที่ h

N คือ จำนวนประชากรทั้งหมด

N_h คือ จำนวนประชากรในชั้นภูมิที่ h

n_h คือ จำนวนกลุ่มตัวอย่างในชั้นภูมิที่ h

เนื่องจากค่ามัชฌิมเลขคณิตของกลุ่มตัวอย่างในแต่ละชั้นภูมิ ($\bar{X}_h$) เป็นค่าซึ่งคำนวณจากกลุ่มตัวอย่างที่มาได้จากการสุ่มตัวอย่างแบบง่าย และเป็นตัวประมาณค่าที่ไม่เอนเอียงของ $\bar{X}_h$ ดังนั้น

$$E(\bar{X}_h) = \bar{X}_h$$

ซึ่งจะทำให้ได้ว่า

$$E(\bar{X}_{st}) = \frac{E(N_1\bar{X}_1+N_2\bar{X}_2+N_3\bar{X}_3+\ldots+N_h\bar{X}_h)}{N}$$

$$= \frac{N_1\bar{X}_1+N_2\bar{X}_2+N_3\bar{X}_3+\ldots+N_h\bar{X}_h}{N}$$

$$= \frac{\sum_{i=1}^{N} X_1+\sum_{i=1}^{N} X_2+\sum_{i=1}^{N} X_3+\ldots+\sum_{i=1}^{N} X_h}{N}$$

$$= \frac{\sum_{h=1}^{L}\sum_{i=1}^{N} X}{N} = \bar{X}$$

เพราะฉนั้น $\bar{X}_{st}$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของ $\bar{X}$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. <u>ค่าความแปรปรวน</u> เราสามารถคำนวณได้จากสูตร (Yamane, 1967)

$$s^2_{st} = \frac{\sum_{h=1}^{L} N_h (s^2_h + (\bar{X}_h - \bar{X}_{st})^2)}{N}$$

โดยที่ $$s^2_h = \frac{\sum_{i=1}^{n_h} (X_{hi} - \bar{X}_h)^2}{n_h - 1}$$

เมื่อ s^2_{st} คือ ค่าความแปรปรวนของกลุ่มตัวอย่าง

s^2_h คือ ค่าความแปรปรวนของกลุ่มตัวอย่างในชั้นภูมิที่ h

จากสูตรการหาค่าความแปรปรวน

$$\begin{aligned}
\sigma^2 &= (1/N) \sum_{h=1}^{L} \sum_{i=1}^{N_h} (X_{hi} - \bar{X})^2 \\
&= (1/N) \sum_{h=1}^{L} \sum_{i=1}^{N_h} [(X_{hi} - \bar{X}_h) + (\bar{X}_h - \bar{X})]^2 \\
&= (1/N)[\sum_{h=1}^{L} \sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2 + \sum_{h=1}^{L} N_h (\bar{X}_h - \bar{X})^2] \\
&= (1/N) \sum_{h=1}^{L} N_h \sigma^2_h + (1/N) \sum_{h=1}^{L} N_h (\bar{X}_h - \bar{X})^2 \\
&= \frac{\sum_{h=1}^{L} N_h (\sigma^2_h + (\bar{X}_h - \bar{X})^2)}{N}
\end{aligned}$$

เนื่องจาก $E(\bar{X}_{st}) = \bar{X}_{st}$

$E(\bar{X}_h) = \bar{X}_h$

$E(s^2) = \sigma^2$

ดังนั้น

$$s^2_{st} = \frac{\sum_{h=1}^{L} N_h (s^2_h + (\bar{X}_h - \bar{X})^2)}{N}$$

จากสูตรการหาค่าความแปรปรวนสามารถจัดรูปได้ใหม่ ดังนี้

$$\sigma^2 = (1/N) \sum_{h=1}^{L} N_h \sigma^2_h + (1/N) \sum_{h=1}^{L} N_h (\bar{X}_h - \bar{X})^2$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เราสามารถนิสูจน์ได้ว่าถ้าผู้วิจัยสามารถจำแนกชั้นภูมิให้ภายในชั้นภูมิมีความเป็นเอกพันธ์ และระหว่างชั้นภูมิมีความเป็นวิวิธพันธ์แล้ว จะทำให้ค่าความคลาดเคลื่อนของค่าประมาณพารามิเตอร์ลดลงหรือกล่าวอีกนัยหนึ่งว่าค่าประมาณที่ได้มีความถูกต้องแม่นยำมากขึ้น ได้ดังนี้

กำหนดให้ $\sigma^2_w = (1/N)\sum_{h=1}^{L} N_h \sigma^2_h$

$\sigma^2_b = (1/N)\sum_{h=1}^{L} N_h (\bar{X}_h - \bar{X})^2$

เมื่อ σ^2_w คือ ค่าความแปรปรวนภายในแต่ละชั้นภูมิของประชากร (within variance)

σ^2_b คือ ค่าความแปรปรวนของประชากรระหว่างชั้นภูมิ (between variance)

ดังนั้น $\sigma^2 = \sigma^2_w + \sigma^2_b$

จากสมการดังกล่าว จะพบว่า ถ้า σ^2_w มีค่าน้อย σ^2_b จะมีค่ามาก นั่นคือ ถ้าประชากรในแต่ละชั้นภูมิถูกจำแนกให้ภายในแต่ละชั้นภูมิมีลักษณะเป็นเอกพันธ์ (Homogeneous) ค่าของ $\sigma^2_w = 0$ ค่าความแปรปรวนที่เกิดขึ้นทั้งหมดจะมาจาก σ^2_b ในทางกลับกัน ถ้าประชากรภายในชั้นภูมิเดียวกันมีความแตกต่างกันมาก (Heterogeneous) ค่าความแปรปรวนที่เกิดขึ้นจะมาจาก σ^2_w การจำแนกให้ภายในแต่ละชั้นภูมิมีความเป็นเอกพันธ์ ($\sigma^2_w = 0$) จะส่งผลให้ค่าความแปรปรวนของตัวประมาณค่า หรือค่าของ $V(\bar{x}_{st})$ มีค่าลดลงด้วยกล่าวอีกนัยหนึ่งว่าค่าประมาณที่ได้มีความถูกต้องแม่นยำมากขึ้น ทั้งนี้เนื่องมาจากในการคำนวณหาค่าความแปรปรวนของตัวประมาณค่า ($V(\bar{x}_{st})$) เราสามารถคำนวณได้จากสูตร

$$V(\bar{x}_{st}) = \sum_{h=1}^{L} \left[\frac{N_h}{N}\right] \cdot \frac{N_h - n_h}{N_h} \cdot \frac{S_h^2}{n_h} \qquad (1)$$

โดยที่ $S_h^2 = \dfrac{\sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2}{N_h - 1}$

เมื่อ $V(\bar{x}_{st})$ คือ ค่าความแปรปรวนของตัวประมาณค่า

S_h^2 คือ ค่าความแปรปรวนของประชากรในชั้นภูมิที่ h

เมื่อพิจารณา ค่าของ S_h^2 จะพบว่าเมื่อค่าของ N_h มีขนาดใหญ่เพียงพอ ค่าของ N_h-1 จะมีค่าเข้าใกล้ N_h ซึ่งทำให้ $S_h^2 \simeq \sigma_h^2$

ดังนั้น $$\sigma_w^2 = (1/N)\sum_{h=1}^{L} N_h.\sigma_h^2 = (1/N)\sum_{h=1}^{L} N_h.S_h^2 \qquad (2)$$

จากสูตร (1) และ (2) เราจะเห็นว่าค่าของ σ_w^2 มีส่วนสัมพันธ์กับค่าของ S_h^2 โดยจะมีค่าแปรผันตามกัน ในขณะเดียวกัน ค่าของ $V(\bar{x}_{st})$ ก็แปรผันตามค่าของ S_h^2 ด้วย จากความสัมพันธ์ดังกล่าวทำให้เราสามารถสรุปได้ว่า ถ้าค่าความแปรปรวนภายในชั้นภูมิ(σ_w^2)มีค่าน้อยลง คือสามารถจำแนกให้ภายในชั้นภูมิมีลักษณะเป็นเอกพันธ์แล้ว ค่าความแปรปรวนของตัวประมาณค่าก็จะมีค่าน้อยลงด้วย นั่นคือ ทำให้ค่าประมาณที่ได้มีความแม่นยำมากยิ่งขึ้น

<u>ตัวแปรจำแนกชั้นภูมิ</u> (Stratified variable)

จากหลักของการสุ่มตัวอย่างแบบแบ่งชั้น คือ การจำแนกประชากรออกเป็นชั้นภูมิย่อย โดยจัดให้สมาชิกภายในแต่ละชั้นภูมิมีความเป็นเอกพันธ์ (Homogeneous) และจัดให้สมาชิกระหว่างชั้นมีความแตกต่างกันหรือวิวิธพันธ์ (Heterogeneous) เราจะพบว่าส่วนที่มีความสำคัญในการจำแนกชั้นประชากรได้ตรงตามวัตถุประสงค์ของการสุ่มแบบแบ่งชั้นก็คือ "ตัวแปรจำแนกชั้นภูมิ" (Stratified variable) ตัวแปรจำแนกชั้นภูมิที่ดีจะทำให้ประสิทธิภาพในการสุ่มตัวอย่างมีมากยิ่งขึ้น เพราะกลุ่มตัวอย่างที่ได้จากการสุ่มจะมาจากทุกส่วนของประชากร การประมาณค่าพารามิเตอร์ก็จะมีความถูกต้องแม่นยำ ดังนั้นในการเลือกตัวแปรจำแนกชั้นภูมิผู้วิจัยจึงควรคำนึงถึงสิ่งต่อไปนี้

1. ตัวแปรจำแนกชั้นภูมิจะต้องจำแนกชั้นประชากรได้สอดคล้องและตรงตามวัตถุประสงค์ที่ผู้วิจัยต้องการศึกษา

2. ตัวแปรจำแนกชั้นภูมิควรเป็นตัวแปรที่มีความสัมพันธ์กันสูงตัวแปรตาม แต่เป็นตัวแปรที่มีความสัมพันธ์กันต่ำกับตัวแปรจำแนกชั้นภูมิตัวอื่น (ในกรณีที่มีตัวแปรจำแนกชั้นหลายตัว)

3. ตัวแปรจำแนกชั้นภูมิที่มีลักษณะเป็นตัวแปรเชิงคุณภาพ เช่น เพศ คณะ สาขาวิชา ภาค จังหวัด เป็นตัวแปรที่มักจะนิยมใช้กันเพราะมีความสะดวกและชัดเจนในการแบ่งชั้นมากกว่าตัวแปรเชิงปริมาณ

4. ในกรณีที่ใช้ตัวแปรเชิงปริมาณในการจำแนกชั้น ควรกำหนดจำนวนชั้นภูมิให้มีความเหมาะสมและสอดคล้องกับวัตถุประสงค์ เพราะการกำหนดจำนวนชั้นที่มากเกินไปอาจทำให้เกิดการสิ้นเปลืองค่าใช้จ่ายในการเก็บข้อมูลโดยไม่จำเป็น นอกจากนี้อาจทำให้เกิดความคลาดเคลื่อนของข้อมูลได้โดยง่าย ในทางตรงกันข้ามถ้าแบ่งจำนวนชั้นน้อยเกินไปก็จะทำให้ประชากรในแต่ละชั้นภูมิมีความหลากหลาย กลุ่มตัวอย่างที่ได้อาจไม่ใช่ตัวแทนที่ดีเท่าที่ควร

<u>วิธีกำหนดขนาดกลุ่มตัวอย่างในแต่ละชั้นภูมิ</u> (Allocation of sample size to strata)

การสุ่มตัวอย่างแบบแบ่งชั้นนี้ นอกจากจะต้องกำหนดขนาดกลุ่มตัวอย่างทั้งหมดแล้วยังต้องพิจารณาถึง ขนาดตัวอย่างในแต่ละชั้นภูมิอีกด้วย ทั้งนี้เพราะจุดมุ่งหมายที่สำคัญอย่างหนึ่งของการสุ่มตัวอย่างแบบแบ่งชั้นก็คือ การลดค่าความคลาดเคลื่อน หรือการเพิ่มความแม่นยำ ถูกต้องของค่าประมาณและการประหยัดค่าใช้จ่ายในการสำรวจ ปัจจัยที่มีผลต่อการกำหนดขนาดกลุ่มตัวอย่างในแต่ละชั้นได้แก่

1. จำนวนสมาชิกทั้งหมดในแต่ละชั้นภูมิ
2. ความแปรปรวนของข้อมูลภายในแต่ละชั้นภูมิ
3. ค่าใช้จ่ายในการเก็บข้อมูลในแต่ละชั้นภูมิ

จากปัจจัยดังกล่าวทำให้วิธีการกำหนดขนาดกลุ่มตัวอย่างย่อยมีความแตกต่างกันออกไปตามความเหมาะสม ซึ่งเราสามารถแบ่งออกเป็น 4 วิธี ดังนี้

1. วิธีกำหนดขนาดกลุ่มตัวอย่างแบบเท่าเทียมกัน (Equal Allocation) วิธีนี้เป็นวิธีที่จำแนกขนาดตัวอย่าง n ออกเป็นส่วนๆ ส่วนละเท่าๆกันในแต่ละชั้นภูมิ หมายความว่าถ้าเราแบ่งประชากรออกเป็น L ชั้นภูมิ ขนาดของกลุ่มตัวอย่างในแต่ละชั้นภูมิจะมีขนาดเท่าๆกัน คือ n/L

นั่นคือ

$$n_h = \frac{n}{L} \quad ; h = 1,2,3...,L$$

โดยที่ $\sum_{h=1}^{L} n_h = n$

เมื่อ n คือ ขนาดของกลุ่มตัวอย่าง

n_h คือ ขนาดของกลุ่มตัวอย่างในชั้นภูมิที่ h

L คือ จำนวนชั้นภูมิ

การกำหนดขนาดตัวอย่างด้วยวิธีนี้เหมาะสำหรับกรณีที่ชั้นภูมิต่างๆมีขนาดเท่ากันหรือใกล้เคียงกัน ไม่เหมาะที่จะใช้กับกรณีที่ชั้นภูมิมีขนาดต่างกันมากนัก

2. วิธีกำหนดขนาดกลุ่มตัวอย่างตามสัดส่วนของชั้นภูมิ (Proportion Allocation) ในกรณีที่ชั้นภูมิมีขนาดแตกต่างกันและไม่อาจถือได้ว่าเท่ากันได้โดยประมาณ ในกรณีเช่นนี้ให้ยึดถือเอาขนาดของชั้นภูมิเป็นเกณฑ์กำหนดน้ำหนัก คือ ชั้นภูมิใดมีขนาดใหญ่กว่าก็ควรจัดสรรให้จำนวนตัวอย่างมากกว่า ชั้นภูมิใดมีขนาดเล็กกว่าก็จัดสรรจำนวนตัวอย่างให้น้อยกว่าโดยคิดเทียบตามหลักของอัตราส่วนซึ่งสามารถเขียนอยู่ในรูปของสูตรได้ดังนี้

$$n_h = \left[\frac{n}{N}\right] \cdot N_h \qquad ; h = 1,2,3\ldots,L$$

เมื่อ n_h คือ ขนาดของกลุ่มตัวอย่างในชั้นภูมิที่ h

N_h คือ จำนวนประชากรในชั้นภูมิที่ h

การกำหนดขนาดกลุ่มตัวอย่างตามสัดส่วนของขนาดของชั้นภูมินี้เป็นวิธีที่นิยมใช้ในทางปฏิบัติ เพราะนอกจากจะสะดวกและรวดเร็วและง่ายต่อการทำความเข้าใจแล้ว การกำหนดขนาดของกลุ่มตัวอย่างโดยวิธีนี้จะมีความแม่นยำสูงมากขึ้นถ้าใช้กับกลุ่มตัวอย่าง n ที่มีขนาดใหญ่พอ

3. วิธีกำหนดขนาดกลุ่มตัวอย่างแบบเหมาะสมหรือแบบอุดมะ (Optimum Allocation) วิธีการกำหนดขนาดของกลุ่มตัวอย่างแบบนี้ต้องคำนึงถึงค่าใช้จ่ายในการสำรวจต่อหน่วยในแต่ละชั้นภูมิด้วย เพราะค่าใช้จ่ายในการสำรวจนั้นโดยปรกติจะแตกต่างกันไปตามชั้นภูมิ เช่น ความยากง่ายในการเข้าถึงหน่วยสำรวจ ระยะทาง ระดับค่าครองชีพในสถานที่ที่ตั้งของหน่วยสำรวจ และอื่นๆ สิ่งเหล่านี้ย่อมส่งผลสะท้อนต่อค่าจ้าง ค่าเดินทาง ซึ่งปัจจัยเหล่านี้เป็นเครื่องชี้ให้เห็นว่าการไม่

คำนึงถึงค่าใช้จ่ายในการสำรวจย่อมส่งผลต่องานและคุณภาพของงานด้วย ดังนั้น วิธีการกำหนดขนาดกลุ่มตัวอย่างตามแนวคิดดังกล่าวจะต้องสอดคล้องกับงบประมาณที่มีอยู่และยังจะต้องเป็นวิธีที่ให้ความแม่นยำสูงสุด หรือ $V(\hat{\bar{y}})$ มีค่าต่ำสุดด้วย จากความสัมพันธ์ดังกล่าวเราสามารถกำหนดขนาดกลุ่มตัวอย่างได้จากสูตร

$$n_h = \frac{N_h.S_h / \sqrt{c_h}}{\sum_{h=1}^{L} (N_h.S_h / \sqrt{c_h})} \cdot n \quad ; h = 1,2,3...,L$$

เมื่อ S_h คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรย่อยในชั้นภูมิที่ h

c_h คือ ค่าใช้จ่ายต่อหน่วยในชั้นภูมิที่ h

เมื่อพิจารณาจากสูตรจะพบว่าค่าของ n_h จะแตกต่างกันไปดังนี้ คือ

3.1 ถ้าชั้นภูมิใดมีขนาด (N_h) ใหญ่กว่าชั้นภูมินั้นก็จะได้รับการจัดสรรให้มีจำนวนตัวอย่างมากกว่า

3.2 ถ้าชั้นภูมิใดมีความแปรปรวนภายในชั้น (Within stratum variance S_h^2) สูงกว่าก็จะได้รับการจัดสรรให้มีจำนวนตัวอย่างที่มากกว่า

3.3 ถ้าชั้นภูมิใดเสียค่าใช้จ่ายต่อหน่วย (c_h) สูงกว่าก็จะได้รับการจัดสรรให้มีจำนวนตัวอย่างที่น้อยกว่า

นอกจากนี้ ในกรณีที่ค่าใช้จ่ายต่อหน่วยในการสำรวจแต่ละชั้นภูมิมีค่าไม่แตกต่างกัน และค่าความแปรปรวนในแต่ละชั้นภูมิก็ไม่แตกต่างกันด้วยแล้ว เราจะพบการกำหนดขนาดตัวอย่างแบบอุดมะนี้ ก็คือการกำหนดขนาดตัวอย่างแบบสัดส่วนนั่นเอง

4. การกำหนดขนาดตัวอย่างแบบนีย์แมน (Neyman Allocation)

จากสูตรการคำนวณหาขนาดกลุ่มตัวอย่างแบบเหมาะสมหรือแบบอุดมะนั้นในสถานการณ์ทางปฏิบัตินั้น บางครั้งค่าใช้จ่ายในการสำรวจต่อหน่วยในชั้นภูมิต่างๆ อาจมีค่าใกล้เคียงกัน หรือไม่แตกต่างกันอย่างชัดเจน ซึ่งในกรณีนี้เราอาจถือว่าค่าใช้จ่ายต่อหน่วยเหล่านั้นอาจเท่ากันได้ ($c_1=c_2=c_3=...=c_L = c$ เมื่อ c เป็นค่าคงที่ใดๆ) เราจะได้ว่า

$$n_h = \frac{N_h \cdot S_h}{\sum_{h=1}^{L} N_h \cdot S_h} \cdot n \quad ; h = 1,2,3...,L$$

เมื่อพิจารณาจากสูตรจะพบว่า ขนาดของกลุ่มตัวอย่างในแต่ละชั้นภูมิจะผันแปรตามขนาดของชั้นภูมิและความแปรปรวนภายในชั้นภูมิ กล่าวคือ ชั้นภูมิใดที่มีขนาดใหญ่กว่าจะได้รับการจัดสรรกลุ่มตัวอย่างที่มากกว่า และชั้นภูมิใดที่มีความแปรปรวนภายในมากกว่าหรือขาดความเป็นเนื้อเดียวกันเราจะต้องจัดสรรตัวอย่างให้มากกว่า

<u>ประโยชน์ของการสุ่มตัวอย่างแบบแบ่งชั้น</u> เราสามารถสรุปได้ดังนี้

1. ง่ายต่อการควบคุมงานสนาม เพราะการปฏิบัติงานสนามอาจทำให้เสร็จสิ้นทีละชั้นภูมิเรื่อยไปจนครบทุกชั้นภูมิ ผู้ควบคุมงานสนามสามารถควบคุมงานได้อย่างใกล้ชิดและสามารถตัดสินใจได้ทันที

2. กลุ่มตัวอย่างรวม (Combine Sample) เป็นตัวแทนที่ดีของประชากรที่เราต้องการศึกษา เพราะจะประกอบไปด้วยหน่วยตัวอย่างที่มาจากทุกส่วนของประชากร

3. ค่าประมาณพารามิเตอร์ที่ได้จากแผนการสุ่มตัวอย่างแบบแบ่งชั้นมีความแม่นยำมากกว่าค่าประมาณที่ได้จากแผนการสุ่มตัวอย่างแบบง่าย นั่นคือ $V(\hat{\theta}_{st}) \leq V(\hat{\theta}_{ran})$ หรือกล่าวอีกนัยหนึ่งว่า วิธีสำรวจแบบแบ่งชั้นภูมิให้ช่วงของการประมาณค่าแคบกว่าวิธีการสุ่มแบบง่าย

4. เนื่องจากการสุ่มตัวอย่างในแต่ละชั้นภูมิเป็นไปได้โดยอิสระทำให้ได้กลุ่มตัวอย่างอิสระ L กลุ่มตามจำนวนชั้นภูมิ ซึ่งประโยชน์ที่จะได้รับอีกประการหนึ่งก็คือ ทำให้เราสามารถเปรียบเทียบค่าประมาณพารามิเตอร์ระหว่างชั้นภูมิได้

<u>ตอนที่ 3 ความคลาดเคลื่อนที่เกิดขึ้นจากการสำรวจ</u> (Component of survey error))

ในการสำรวจด้วยตัวอย่าง จะมีความคลาดเคลื่อนที่อาจเกิดขึ้นได้อยู่ 2 ลักษณะใหญ่ๆ คือ (อภิชาติ พงษ์ศรีหดุลชัย, 2530; ศิริชัย กาญจนวาสี, 2533)

1. ความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่าง (Sampling Error) เป็นความคลาดเคลื่อนที่เกิดขึ้น เพราะไม่ได้ทำการสังเกตหรือวัดค่าจากทุกหน่วยของประชากร ดังนั้น

ค่าที่ประมาณได้จากตัวอย่างจึงมักจะแตกต่างไปจากค่าที่แท้จริงของประชากรเสมอไม่มากก็น้อย อย่างไรก็ตาม โดยอาศัยทฤษฎีการสุ่มตัวอย่าง เราจะสามารถประมาณความคลาดเคลื่อนดังกล่าวนี้ได้ แม้ว่าตัวอย่างที่สำรวจจะเป็นเพียงตัวอย่างเดียวจากโอกาสที่เป็นไปได้ทั้งหมดหลายๆแบบก็ตาม ซึ่งค่าที่ใช้วัดความคลาดเคลื่อนนี้ คือ ค่าความแปรปรวนจากการสุ่ม (Sampling Variance)

2. ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง (Non-sampling Error) เป็นความคลาดเคลื่อนที่เกิดขึ้นได้เสมอ แม้แต่การสำรวจหรือเก็บข้อมูลจากทุกหน่วยของประชากรก็ตาม ทั้งนี้อาจเกิดจากหลายสาเหตุด้วยกัน ทั้งในภาคสนามและสำนักงาน เช่น เครื่องมือขาดความเที่ยงตรง ผู้เก็บรวบรวมข้อมูลขาดคุณสมบัติที่ดี การวัดค่าผิดพลาดได้ข้อมูลหรือคำตอบไม่ถูกต้อง (Response Error) การบันทึกข้อมูลหรือลงรหัสผิดพลาด การลืมวัดค่าบางหน่วย รวมทั้งการใช้กรอบที่มีการนับจดตกหล่น (Omisssion) หรือ นับซ้ำ (Duppli-cation) เป็นต้น ซึ่งปรกติแล้วเราจะไม่สามารถวัดค่าความคลาดเคลื่อนชนิดนี้ได้

การวัดคุณภาพของตัวประมาณค่า นอกจากการวัดความแม่นยำ (Precision) ซึ่งใช้ค่าความแปรปรวนจากการสุ่ม ($V(\hat{\theta})$)เป็นตัววัดแล้ว ยังมีอีกค่าหนึ่ง เรียกว่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error , MSE) ซึ่งเป็นค่าที่ใช้วัดความเที่ยงตรง หรือความถูกต้อง (Accuracy) ของตัวประมาณค่า และมีคำนิยามดังนี้

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

เมื่อ θ คือ ค่าที่แท้จริงของประชากร

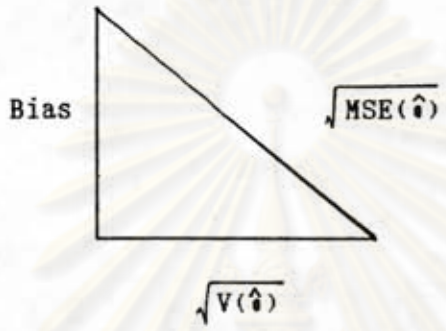
$\hat{\theta}$ คือ ตัวประมาณค่าของ θ

จากคำนิยามดังกล่าวเราสามารถเขียนได้ใหม่ ดังนี้

$$\begin{aligned}
MSE(\hat{\theta}) &= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\
&= E[(\hat{\theta}-E(\hat{\theta}))^2+(E(\hat{\theta})-\theta)^2+2(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)] \\
&= E[(\hat{\theta}-E(\hat{\theta}))^2+E(E(\hat{\theta})-\theta)^2+2E((\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)] \\
&= V(\hat{\theta})+(E(\hat{\theta})-\theta)^2+2(E(\hat{\theta})-\theta)E(\hat{\theta}-E(\hat{\theta})) \\
&= V(\hat{\theta}) + Bias^2 + 2(E(\hat{\theta})-\theta)(E(\hat{\theta})-E(\hat{\theta})) \\
&= V(\hat{\theta}) + Bias^2
\end{aligned}$$

โดยที่ Bias = $(E(\hat{\theta})-\theta)$ หมายถึง ค่าที่เอนเอียงไปจากค่าที่แท้จริงของประชากร ซึ่งถ้าเขียนเป็นรูปสามเหลี่ยมมุมฉากจะได้ความสัมพันธ์ระหว่าง $MSE(\hat{\theta})$, $V(\hat{\theta})$, Bias ดังนี้

แผนภาพที่ 4 ความสัมพันธ์ระหว่างค่าความแปรปรวน ค่าความเอนเอียง และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง



ถ้าหาก $\hat{\theta}$ เป็นตัวประมาณค่าที่ไม่เอนเอียงแล้ว $(E(\hat{\theta}) = \theta)$ จะทำให้ Bias เท่ากับศูนย์ ดังนั้น จะได้ $MSE(\hat{\theta}) = V(\hat{\theta})$ นั่นคือ ความเที่ยงตรงก็คือ ความแม่นยำนั่นเอง

ตอนที่ 4 การประมาณค่า (Estimation)

จุดมุ่งหมายประการสำคัญในการสุ่มกลุ่มตัวอย่างมาจากประชากรก็คือ การประมาณค่าของพารามิเตอร์อันเป็นลักษณะของประชากรที่ผู้วิจัยสนใจ ในการประมาณค่าของประชากรกระทำได้ 2 แบบ คือ การประมาณค่าแบบจุด (Point estimstion) และการประมาณค่าแบบช่วง (Interval estimation) วิธีการประมาณค่าแบบจุดเป็นการประมาณค่าของประชากรที่สนใจศึกษาด้วยค่าเพียงค่าเดียวเท่านั้น เช่น ใช้ $\bar{x}$ ประมาณค่า μ หรือ s^2 ประมาณค่า σ^2 เป็นต้น ส่วนวิธีการประมาณค่าแบบช่วงเป็นการประมาณค่าพารามิเตอร์ของประชากรที่สนใจศึกษาด้วยช่วงค่าช่วงหนึ่ง ซึ่งมีคุณสมบัติว่าค่าของประชากรที่แท้จริงจะตกอยู่ในช่วงค่าที่ประมาณนี้ ด้วยความเชื่อมั่นระดับหนึ่งโดยจะต้องอาศัยการประมาณค่าแบบจุดและการแจกแจงความน่าจะเป็นของตัวประมาณเป็นพื้นฐานในการคำนวณ การประมาณค่าทั้งสองแบบจะเหมาะสมกับกรณีการใช้งานที่ต่างกัน ในกรณีที่มีตัวประมาณค่า (Estimator) ซึ่งสามารถนำมาประมาณค่าพารามิเตอร์ได้หลายวิธี

ได้มีการกำหนดคุณสมบัติของวิธีการประมาณค่าที่ควรมีคุณสมบัติ 4 ประการดังต่อไปนี้ (Hay, 1963; Yamane, 1967)

1. ความไม่เอนเอียง (Unbiasness)

ถ้า $\hat{\theta}$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของพารามิเตอร์ θ แล้วจะได้ว่า

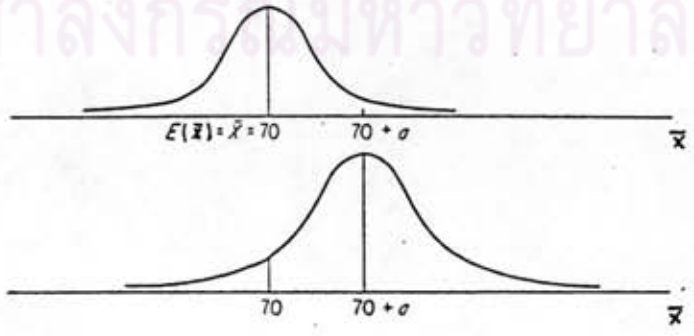
$$E(\hat{\theta}) = \theta$$

นั่นคือ ค่าคาดหวัง (Expected value) ของตัวประมาณค่า $\hat{\theta}$ มีค่าเท่ากับค่าของพารามิเตอร์ตัวที่ต้องการประมาณหรือกล่าวอีกนัยหนึ่งว่า ตัวประมาณค่านั้นไม่เอนเอียง ถ้า Expectation ของการแจกแจงของ $\hat{\theta}$ มีค่าเท่ากับพารามิเตอร์ θ และตัวประมาณค่าที่ไม่เอนเอียงนั้นมีคุณสมบัติที่ดีอยู่ว่า ถ้ามีชุดของค่าประมาณที่ไม่เอนเอียงที่เป็นอิสระต่อกันอยู่แล้วค่าเฉลี่ยเหล่านั้นย่อมไม่เอนเอียงด้วย และในทางตรงกันข้ามค่าเฉลี่ยของค่าประมาณที่เอนเอียงย่อมเอนเอียงด้วยไม่ว่าจะเฉลี่ยมากี่ค่าก็ตาม

ส่วนค่าเอนเอียง (Bias) ของตัวประมาณ $\hat{\theta}$ คือ ค่าความแตกต่างระหว่างค่าที่คาดหวังของ $\hat{\theta}$ กับค่าพารามิเตอร์ θ กล่าวคือ

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

แผนภาพที่ 5 การแจกแจงของตัวประมาณค่าที่ไม่เอนเอียง และตัวประมาณค่าที่เอนเอียง



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

I17194736

2. ความคงเส้นคงวา (Consistency)

การประมาณค่าพารามิเตอร์ θ ด้วย $\hat{\theta}$ เมื่อขนาดของกลุ่มตัวอย่างใหญ่ขึ้น ตัวประมาณ $\hat{\theta}$ มีค่าเข้าใกล้พารามิเตอร์ θ มากขึ้นด้วย ซึ่งสามารถเขียนอยู่ในรูปประโยคทางคณิตศาสตร์ได้ว่า

$$\lim_{n\to\infty} p(\left|\hat{\theta}_n - \theta\right| < \varepsilon) = 1$$

เมื่อ ε คือ ค่าความคลาดเคลื่อนที่ยอมให้เกิดขึ้นได้ในการหาค่าพารามิเตอร์ด้วยตัวประมาณค่า ($\hat{\theta}$) โดยที่ P คือ ความน่าจะเป็นหรือโอกาสที่จะเกิดเหตุการณ์นั้นขึ้น

เราจะเรียก $\hat{\theta}$ ว่า เป็นตัวประมาณค่าที่มีความคงเส้นคงวาของ θ ซึ่งหมายความว่า ความน่าจะเป็นที่ค่า $\hat{\theta}$ จะประมาณเข้าใกล้ θ มีค่าสูงขึ้น เมื่อขนาดกลุ่มตัวอย่าง (n) มีค่าเพิ่มมากขึ้นๆ คือมีค่าความน่าจะเป็นเข้าใกล้ 1

3. ความมีประสิทธิภาพ (Efficiency)

หมายถึง การนิจารณาถึงความถูกต้องแม่นยำในการประมาณค่าของตัวประมาณค่าหนึ่งๆ เกณฑ์ที่ใช้นิจารณาความมีประสิทธิภาพของตัวประมาณค่า คือ ความแปรปรวนที่เปรียบเทียบกับกลุ่มของตัวประมาณค่าที่ไม่ลำเอียงด้วยกัน

โดยทั่วไปจะนิยามประสิทธิภาพของตัวประมาณค่าใดๆว่าเป็นอัตราส่วนระหว่างค่าความแปรปรวนของตัวประมาณค่าที่มีประสิทธิภาพนั้นกับค่าความแปรปรวนของตัวประมาณค่าตัวอื่นๆ กล่าวคือ ถ้าค่าความแปรปรวนของ $\hat{\theta}_i$ หรือ $Var(\hat{\theta}_i)$ น้อยกว่าค่าความแปรปรวนของ $\hat{\theta}_j$ หรือ $Var(\hat{\theta}_j)$ เมื่อทั้ง $\hat{\theta}_i$ และ $\hat{\theta}_j$ เป็นตัวประมาณค่าที่ไม่ลำเอียงแล้วจะได้ว่า

เมื่อ $$E_r = \frac{Var(\hat{\theta}_i)}{Var(\hat{\theta}_j)} \quad ; \quad 0 < E_r < 1$$

ดังนั้นเมื่อค่า E_r ที่ได้จากการคำนวณน้อยกว่า 1 แสดงว่า $\hat{\theta}_i$ เมื่อเทียบกับ $\hat{\theta}_j$ แล้ว $\hat{\theta}_i$ มีประสิทธิภาพมากกว่า (Absolutely efficient) เพราะว่าความแปรปรวนของ $\hat{\theta}_i$ เล็กกว่าความแปรปรวนของ $\hat{\theta}_j$

ในการนิจารณาคุณภานของตัวประมาณค่านารามิเตอร์ $\hat{\theta}$ ลักษณะที่น่าจะนิจารณาอีกลักษณะหนึ่งก็คือ ค่าต่างๆของ $\hat{\theta}$ มีความแตกต่างจากค่าที่ต้องการประมาณเพียงใด ทั้งนี้เพราะถ้าค่าส่วนใหญ่ของ $\hat{\theta}$ มีความแตกต่างจากค่า θ น้อย ก็แสดงว่าโอกาสที่ค่าประมาณจะตกอยู่ใกล้ค่าของ θ มาก ตัวประมาณค่าที่มีลักษณะเช่นนี้จะดีกว่าตัวประมาณค่าที่มีค่าเป็นไปได้กระจายออกจากค่าที่ต้องการประมาณมาก ดังนั้นเกณฑ์ที่จะใช้เปรียบเทียบคุณภาพลักษณะนี้คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error) แต่ในทางปฏิบัติเราไม่สามารถที่จะหาตัวประมาณค่าที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสำหรับทุกๆค่าของ θ ได้ จากการนิจารณาตัวประมาณค่าที่ไม่เอนเอียง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจะเท่ากับค่าความแปรปรวน จึงสามารถหาตัวประมาณที่ดีที่สุดในกลุ่มตัวประมาณที่ไม่เอนเอียง คือ ตัวประมาณที่มีค่าความแปรปรวนต่ำสุด

4. ความพอเพียง (Sufficiency)

ตัวประมาณค่า $\hat{\theta}$ จะเป็นตัวประมาณค่าที่มีความพอเพียง ถ้าตัวประมาณค่าให้สารสนเทศที่ก่อเกิดประโยชน์ได้ทั้งหมดที่ต้องการเกี่ยวกับนารามิเตอร์ที่ต้องการประมาณ เช่น $\bar{X}$ เป็นตัวประมาณที่พอเพียงของ μ ก็หมายความว่าไม่มีตัวประมาณค่าของ μ ตัวอื่น เช่น ฐานนิยม (Mode) มัธยฐาน (Median) ที่จะสามารถให้สารสนเทศเกี่ยวกับ μ เพิ่มขึ้นได้อีกดีไปกว่า $\bar{X}$

ตอนที่ 5 การแจกแจงค่าสถิติจากการสุ่มตัวอย่าง (Sampling Distribution)

ในการดำเนินการวิจัย ผู้วิจัยมักจะศึกษากับลักษณะของกลุ่มตัวอย่างมากกว่าที่จะทำการศึกษากับประชากรโดยตรง เพราะข้อจำกัดต่างๆทางทรัพยากร ดังนั้นผู้วิจัยจึงต้องอาศัยเทคนิคการนำค่าสถิติที่ได้จากกลุ่มตัวอย่างไปช่วยในการประมาณค่าหรือสรุปผลของประชากร ผู้วิจัยส่วนใหญ่จึงศึกษาความสัมพันธ์ระหว่างค่าที่ได้จากกลุ่มตัวอย่างกับค่าพารามิเตอร์ของประชากรเพื่อเป็นประโยชน์ในการสรุปอ้างอิงอันจะนำไปสู่การตัดสินใจอย่างใดอย่างหนึ่ง ในการศึกษาข้อมูลจากกลุ่มตัวอย่างนั้นเมื่อทำการสุ่มตัวอย่างขนาด n จากประชากรหนึ่งซึ่งมีขนาด N สามารถมีกลุ่มตัวอย่างที่เป็นไปได้ทั้งหมด ${}_{N}C_{n}$ กลุ่มตัวอย่าง (All possible samples of size n) ในแต่ละกลุ่มตัวอย่างสามารถคำนวนหาค่าสถิติต่างๆ เช่น ค่าเฉลี่ย ค่าสัดส่วน ค่าส่วนเบี่ยงเบนมาตรฐาน ฯลฯ ซึ่งอาจจะมีค่าแตกต่างกันไปในแต่ละกลุ่มตัวอย่าง ถ้านำค่าสถิติเหล่านั้นมาแจกแจงความถี่จะได้รูปแบบการแจกแจงที่เรียกว่า การแจกแจงการสุ่มตัวอย่าง(ศิริชัย กาญจนวาสี, 2533)

ซึ่งการแจกแจงการสุ่มตัวอย่างมีลักษณะแตกต่างกันขึ้นอยู่กับค่าสถิติที่นำมาแจกแจง เช่น เมื่อพิจารณาค่าเฉลี่ยแต่ละค่าของตัวอย่างขนาด n ที่สุ่มได้จากประชากร N ค่าของ $\bar{X}$ ย่อมขึ้นอยู่กับกลุ่มตัวอย่างที่ได้ $\bar{x}$ ซึ่งเป็นตัวแปรสุ่มการแจกแจงของ $\bar{X}$ เรียกว่า การแจกแจงของค่าเฉลี่ยที่ได้จากการสุ่มตัวอย่าง

ตามทฤษฎี เซ็นทรัล-ลิมิต (Central-limit Theorem) กล่าวว่า ถ้ามีประชากรกลุ่มหนึ่งซึ่งมีค่าเฉลี่ยเป็น μ และมีความแปรปรวนจำกัด (finite variance) เป็น σ^2 แล้วการแจกแจงการสุ่มตัวอย่างของ $\bar{X}$ (Sampling distribution of mean) ของกลุ่มตัวอย่างขนาด n ที่สุ่มมาจากประชากรนั้นจะเข้าใกล้การแจกแจงปรกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2/n เมื่อ n เพิ่มขึ้นไม่จำกัด ($n \to \infty$) (ผจงจิต อินทรสุวรรณ, 2528) อาศัยทฤษฎีทางสถิติลักษณะการแจกแจงการสุ่มของค่าสถิติ $\bar{X}$ ที่สุ่มมาจากประชากรขนาดเท่ากับ N เมื่อทุกการสุ่มกลุ่มตัวอย่างขนาด n จะได้การแจกแจงที่มีลักษณะใกล้เคียงกับโค้งปรกติ โดยมีคุณสมบัติดังนี้

1. ค่าเฉลี่ยของการแจกแจงการสุ่มของค่าสถิติ $\bar{X}$ ($E(\bar{X})$) เท่ากับค่าเฉลี่ยของประชากร ($E(\bar{X}) = \mu$)

2. ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงการสุ่มของค่าสถิติ $\bar{X}$ เรียกว่า ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย (Standard error of mean หรือ SE) ใช้สัญลักษณ์ว่า $\sigma_{\bar{x}}$ มีค่าเท่ากับ ส่วนเบี่ยงเบนมาตรฐานของประชากรหารด้วยกรณฑ์ที่สองของขนาดของกลุ่มตัวอย่าง นั่นคือ ถ้าขนาดตัวอย่าง (n) มีขนาดใหญ่ $\sigma_{\bar{x}} = \sigma/\sqrt{N}$ แต่ถ้ากลุ่มตัวอย่างขนาดเล็ก

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \cdot \sigma/\sqrt{N}$$ (อุทุมพร ทองอุไทย, 2523)

ดังนั้นการสุ่มตัวอย่างมาจากประชากรแล้วใช้ข้อมูลจากการสุ่มตัวอย่างเพื่อคำนวณค่าสถิติหรือ ตัวประมาณค่าเพื่อประมาณค่าพารามิเตอร์ รูปร่างของการแจกแจงของค่าเฉลี่ยของกลุ่มตัวอย่างถ้าประชากรมีการแจกแจงเป็นปรกติ กลุ่มตัวอย่างที่สุ่มมาจากประชากรนี้ย่อมประกอบด้วยคะแนนที่มีการแจกแจงเป็นปรกติด้วย แต่ถ้าประชากรไม่เป็นปรกติทำให้สามารถคาดการณ์เกี่ยวกับการแจกแจงโดยอาศัยทฤษฎีเซ็นทรัล-ลิมิต ทฤษฎีนี้เป็นทฤษฎีที่สำคัญที่สุดในทฤษฎีทั้งหลายของความน่าจะเป็นและสถิติ ซึ่งจะช่วยบอกเกี่ยวกับการแจกแจงโดยประมาณของค่าเฉลี่ยของกลุ่มตัวอย่าง $\bar{X}$ ได้

ดังจะเห็นได้ว่าในการประมาณค่าหรือการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ที่สนใจมีความจำเป็นที่ผู้วิจัยจะต้องทราบลักษณะการแจกแจงของค่าสถิติต่างๆ เพื่อเป็นประโยชน์ในการ

เลือกใช้สถิติได้เหมาะสม (กนกทิพย์ พัฒนาพันธ์, 2529)

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

ในการศึกษางานวิจัยที่เกี่ยวกับการพัฒนารูปแบบและวิธีการสุ่มตัวอย่างนั้น พบว่ามีผู้ที่สนใจและทำการศึกษาเกี่ยวกับรูปแบบและวิธีการสุ่มตัวอย่างต่างๆกัน ดังนี้

ผกามาศ สิงห์สง่า (2523) ได้ศึกษาแผนแบบการสุ่มตัวอย่างแบบแบ่งชั้น 3 แบบ คือ แผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นทางเดียวโดยสุ่มตัวอย่างมาชั้นละ 1 หน่วย แผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นสองทางโดยใช้ตัวอย่างจำนวนน้อย และแผนแบบการสุ่มตัวอย่างแบ่งชั้นสองทางโดยใช้จำนวนแบบมีระบบโดยให้ความน่าจะเป็นของการสุ่มหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ (ในแต่ละชั้นภูมิสุ่มตัวอย่างแบบมีระบบตามจำนวนที่ต้องการจากการคำนวณหาขนาดของกลุ่มตัวอย่างในแต่ละชั้น) ศึกษาเปรียบเทียบประสิทธิภาพว่าแผนแบบใดจะให้ค่าประมาณที่มีความเชื่อถือได้มากกว่ากัน โดยใช้ข้อมูลจำนวนครัวเรือนส่วนบุคคล ของกองสำรวจประชากร จำแนกตามขนาดของครัวเรือนและอำเภอของกรุงเทพมหานคร จากรายงานสำมะโนประชากรและเคหะ พ.ศ. 2513 ของหน่วยสำมะโนประชากรและเคหะ กองสำรวจประชากร สำนักงานสถิติแห่งชาติ พบว่า การใช้แผนการสุ่มตัวอย่างแบบแบ่งชั้นสองทางแบบมีระบบให้ค่าประมาณที่ใกล้เคียงกับค่าจริงมากที่สุด และมีความแปรปรวนน้อยที่สุด ส่วนแผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นทางเดียวโดยสุ่มตัวอย่างมาชั้นละ 1 หน่วยให้ค่าประมาณที่มีความแปรปรวนมากที่สุด

สุรพล ปชานวนิช (2529) ได้ทำการทดสอบสมรรถภาพของการสุ่มตัวอย่างในงานวิจัยทางสังคมศาสตร์ โดยศึกษาเปรียบเทียบการสุ่มตัวอย่างแบบไม่อาศัยทฤษฎีความน่าจะเป็น กับการสุ่มตัวอย่างแบบอาศัยหลักความน่าจะเป็น ซึ่งวิธีการสุ่มตัวอย่างที่ใช้ทดสอบ คือ การสุ่มตัวอย่างแบบเจาะจง การสุ่มตัวอย่างแบบง่าย การสุ่มตัวอย่างแบบมีระบบ การสุ่มตัวอย่างแบบแบ่งชั้น การสุ่มตัวอย่างแบบยกกลุ่ม และการสุ่มตัวอย่างแบบแบ่งชั้นตามคุณลักษณะของตัวแปรอิสระ 2 ตัว ประชากรที่ใช้เป็นข้าราชการพลเรือนระดับกรมหน่วยงานหนึ่ง จำนวน 7,057 คน ผลการศึกษาพบว่า การสุ่มตัวอย่างแบบเจาะจง ซึ่งเป็นตัวแทนของการสุ่มตัวอย่างที่ไม่อาศัยทฤษฎีความน่าจะเป็น มีประสิทธิภาพต่ำที่สุด สำหรับการสุ่มตัวอย่างที่อาศัยความน่าจะเป็นด้วยกัน การสุ่มตัวอย่าง

แบบยกกลุ่มจะมีประสิทธิภาพต่ำที่สุด และการสุ่มตัวอย่างแบบแบ่งชั้นให้ค่าประมาณใกล้เคียงกับพารามิเตอร์สม่ำเสมอที่สุด

นิเวศน์ คำรัตน์ (2534) ศึกษาการเปรียบเทียบค่าประมาณมัชฌิมเลขคณิตของประชากรจากกลุ่มตัวอย่างที่ได้จากการสุ่มแบบต่างๆ ได้แก่ การสุ่มตัวอย่างแบบง่าย การสุ่มตัวอย่างแบบมีระบบ การสุ่มตัวอย่างแบบแบ่งชั้น การสุ่มตัวอย่างแบบสองขั้น และการสุ่มตัวอย่างแบบสามขั้น โดยใช้ขนาดของกลุ่มตัวอย่างที่แตกต่างกัน 3 ขนาด คือ จำนวน 1,095 , 1,555 และ 2,636 คน ประชากรที่ใช้ในการวิจัย คือ ข้าราชการครู ในเขตการศึกษา 1 จนถึงเขตการศึกษา 12 กรมสามัญศึกษา กระทรวงศึกษาธิการ ปีการศึกษา 2531 จำนวน 78,035 คน ดำเนินการสุ่มตัวอย่างในแต่ละวิธีและแต่ละขนาดของกลุ่มตัวอย่างโดยใช้คอมพิวเตอร์กระทำการสุ่มซ้ำ 1,000 ครั้ง เปรียบเทียบค่าประมาณมัชฌิมเลขคณิตของอายุของประชากรตามคุณสมบัติของตัวประมาณค่าที่ดี 3 ด้าน คือความไม่เอนเอียง ความคงเส้นคงวา ความมีประสิทธิภาพ พบว่า วิธีการสุ่มตัวอย่างแบบสองขั้นให้ค่าประมาณมัชฌิมเลขคณิตของประชากรที่มีความเอนเอียง ที่ขนาดของกลุ่มตัวอย่าง จำนวน 1,095 คน วิธีการสุ่มตัวอย่างทั้ง 5 วิธีมีความคงเส้นคงวา และวิธีการสุ่มแบบแบ่งชั้นให้ค่าประมาณมัชฌิมเลขคณิตของประชากรมีประสิทธิภาพมากที่สุด

สมชัย วงษ์นายะ (2534)ได้ศึกษาเปรียบเทียบค่าประมาณพารามิเตอร์จากแบบแผนการสุ่มตัวอย่างต่างแบบ โดยใช้ข้อมูลจากคะแนนผลสัมฤทธิ์วิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ปีการศึกษา 2531 จำนวน 7,335 คน ซึ่งใช้การสุ่มตัวอย่างทั้งหมด 7 วิธี ได้แก่ การสุ่มตัวอย่างแบบง่าย การสุ่มตัวอย่างแบบมีระบบ การสุ่มตัวอย่างแบบแบ่งชั้นที่ใช้ตัวแปรจำแนกชั้นต่างกัน คือ ขนาดโรงเรียน คุณภาพของโรงเรียน อำเภอ และการสุ่มตัวอย่างแบบแบ่งชั้น 2 ระยะโดยใช้ตัวแปรจำแนกชั้นอำเภอ-ขนาดโรงเรียน และตัวแปรจำแนกชั้นอำเภอ-คุณภาพของโรงเรียน ด้วยขนาดกลุ่มตัวอย่าง 3 ขนาด คือ ขนาดกลุ่มตัวอย่างที่มีระดับความเชื่อมั่น 90%, 95% และ 99% (424 คน,588 คน และ 962 คน) ดำเนินการสุ่มตัวอย่างแต่วิธีโดยใช้คอมพิวเตอร์ในการสุ่มซ้ำ 1,000 ครั้ง เปรียบเทียบค่าประมาณมัชฌิมเลขคณิตและค่าประมาณความแปรปรวนผลการศึกษาพบว่า วิธีการสุ่มตัวอย่างแบบแบ่งชั้นที่ใช้ขนาดโรงเรียนเป็นตัวจำแนกชั้น ด้วยขนาดกลุ่มตัวอย่างที่มีความเชื่อมั่น 95% ให้ค่าเฉลี่ยของค่าประมาณมัชฌิมเลขคณิต($\bar{y}$) ใกล้เคียงกับค่าพารามิเตอร์(μ) มากที่สุด และวิธีการสุ่มแบบแบ่งชั้นที่ใช้คุณภาพของโรงเรียนเป็นตัวจำแนกชั้น ด้วยขนาดกลุ่มตัวอย่างที่มีความ

เชื่อมั่น 90% ให้ค่าเฉลี่ยของค่าประมาณความแปรปรวน ($\bar{S}^2$) ใกล้เคียงกับค่าพารามิเตอร์ (σ^2) มากที่สุด นอกจากนั้นบ่าวิธีการสุ่มตัวอย่างแบบเป็นระบบ ด้วยขนาดกลุ่มตัวอย่างที่มีความเชื่อมั่น 99% ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย