

การลุ่มตัวอย่างเพื่อการวิจัยทางสังคมศาสตร์

โดย เชวง เรืองสุวรรณ

การวิจัยเป็นกรรมวิธีเชิงวิทยาศาสตร์ที่ได้รับการพัฒนาปรับปรุงขึ้น เพื่อใช้เป็นเครื่องมือในการเสาะแสวงหาความรู้และความเข้าใจในปรากฏการณ์ต่าง ๆ ทั้งในด้านวิทยาศาสตร์สังคม และวิทยาศาสตร์ธรรมชาติ ความรู้ที่ได้รับจากการวิจัยนั้นจะเป็นในสาขาวิชาการใดก็ตาม แม้จะเป็นเพียงความรู้และความเข้าใจขั้นมูลฐาน หากว่าได้ศึกษาค้นคว้ามาด้วยความระมัดระวัง และเสนอให้เป็นที่ประจักษ์ในวงวิชาการด้วยความสุจริตใจแล้ว ย่อมจะได้รับความนับถือและยอมรับว่าเป็นความรู้ใหม่ที่เพิ่มขึ้น เพราะอย่างน้อยที่สุด ความพยายามในการเสาะแสวงหาให้ได้นั้น และการเสนอให้ทราบซึ่งความรู้ดังกล่าว ย่อมทำให้สาขาวิชาการนั้นเพิ่มความสามารถในการที่จะพยากรณ์และเข้าใจความเป็นไปของธรรมชาติหรือปรากฏการณ์สังคม การที่จะทำให้ผลการวิจัยได้รับการยอมรับนับถือดังกล่าว จำเป็นอย่างยิ่งที่นักวิจัยจะต้องได้รับการฝึกฝนอบรมมาเป็นอย่างดี ทั้งในสาขาวิชาการ และในระเบียบวิธีการวิจัย นอกจากนี้ การยอมให้ความรู้ที่ได้มาจากการวิจัยของตนแสดงปรากฏการณ์ของตนเองแทนการแสดงออกตามความคิดเห็นที่ผู้วิจัยคิดว่ามันควรจะเป็น เป็นหลักการอันสำคัญที่นักวิจัยพึงยึดถือในการดำเนินงาน ปกติวิสัยของมนุษย์มักจะมีความลำเอียงเป็นส่วนตัว นักวิจัยจึงพยายามหลีกเลี่ยงที่จะให้ความคิดเห็นส่วนตัวเข้าไปปะปนยุ่งเกี่ยวกับขอบข่ายการวิจัย ทั้งในขั้นการเก็บรวบรวม การวิเคราะห์ และการเสนอข้อมูล

การวิจัยทางสังคมศาสตร์เป็นระเบียบวิธีการศึกษาที่มีระบบ เพื่อให้ได้มาซึ่งความรู้และความเข้าใจใหม่ ๆ เกี่ยวกับความเป็นไปทางสังคม ปรากฏการณ์สังคมเป็นสิ่งที่มนุษย์ปรารถนาจะเรียนรู้ เพื่อทำความเข้าใจในสังคมที่ตนเป็นส่วนหนึ่ง เพื่อหาทางแก้ปัญหาสังคม เพื่อปรับปรุงตนเองและสังคม เพื่อจัดระบบสังคมให้ดียิ่งขึ้น อันจะเป็นเหตุให้สังคมได้ดำรงอยู่ด้วยความสงบสุข ซึ่งเป็นยอดปรารถนาของมนุษย์ทุกรูปทุกนาม แม้

การวิจัย จะช่วยขยายขอบเขตความรู้ความเข้าใจของมนุษย์เกี่ยวกับสังคมมากขึ้น ตามลำดับของกาลเวลาที่ผ่านไป แต่ก็ปรากฏว่า มนุษย์ยังคงประสบกับความยุ่งยากเกี่ยวกับสังคมที่ตนดำรงอยู่นานาประการ การรวมกันอยู่ของมนุษย์ทำให้เกิดปรากฏการณ์ทางสังคมอันเป็นปัญหาที่ยุ่งยากซับซ้อนนานาชนิด การหาคำตอบต่อปัญหาสังคมนั้น เมื่อพิจารณาจะเห็นว่าลึกซึ้งและสลับซับซ้อนมากกว่าการมุ่งหาคำตอบต่อปัญหาในปรากฏการณ์ทางธรรมชาติ ทั้งนี้เป็นเพราะความรู้และทฤษฎีต่าง ๆ ในทางวิทยาศาสตร์ธรรมชาติได้วางอยู่บนรากฐานอันแน่นอนและมั่นคงกว่า นอกจากนั้น การวิจัยทางวิทยาศาสตร์ธรรมชาติได้พัฒนาอุปกรณ์และเครื่องมือต่าง ๆ ที่สามารถจะควบคุมหรือศึกษาปัจจัยหรือตัวแปรต่าง ๆ ได้ดี ส่วนสภาพการณ์ในทางสังคมนั้น มีปัจจัยอันเป็นสาเหตุต่าง ๆ อยู่เป็นจำนวนมาก ปัจจัยเหล่านี้ย่อมทำให้เกิดผลทางสังคมที่ผันแปรไม่แน่นอนอีกมาก ปัจจัยและผลเกี่ยวเนื่องระหว่างปัจจัยเหล่านี้ ทำให้มนุษย์ไม่สามารถจะสรรหามาตรการที่แน่นอนใด ๆ มาจัดและทำนายปรากฏการณ์สังคมได้อย่างถูกต้องและแน่นอน

แม้ปรากฏการณ์ทางสังคมจะยุ่งยากสลับซับซ้อน นักสังคมวิทยาก็พยายามค้นคว้าศึกษาสภาพการณ์ต่าง ๆ ของสังคม พยายามเก็บรวบรวมและสะสมข้อมูล ปรับปรุงหลักเกณฑ์และระเบียบวิชาการต่าง ๆ เพื่อนำไปไขแก้ไขปัญหาสังคมอยู่ตลอดเวลา พยายามพัฒนาอุปกรณ์และเครื่องมือเครื่องใช้ในการศึกษาและปรับปรุงมาตรการที่มีอยู่ เพื่อเพิ่มความถูกต้องและความสามารถในการพยากรณ์ คาดคะเนความเป็นไปในสังคม และทำความเข้าใจในปรากฏการณ์ทางสังคมให้ดียิ่งขึ้นอยู่ตลอดมา สภาวะทางเศรษฐกิจและสังคม เช่น รายได้ รายจ่าย สถานภาพการทำงาน สุขภาพอนามัย ที่อยู่อาศัย การเคลื่อนย้ายของประชากร ขนาดของครอบครัว การเกิด การตาย ความเชื่อ ความคิดเห็น ทรรศนะในสิ่งต่าง ๆ การปกครองและคุณลักษณะอื่น ๆ อีกนับไม่ถ้วน จึงกลายเป็นจุดสนใจที่สำคัญของการสำรวจวิจัย ยิ่งไปกว่านั้น ความคิด ความรู้สึก การกระทำ การคาดการณ์ล่วงหน้า การซื้อ การอ่าน การฟัง การเมือง การพักผ่อนหย่อนใจและเกือบจะทุกสาขาของกิจกรรมมนุษย์ ต่างก็ได้รับความสนใจและมีผู้ทำการศึกษาค้นคว้าอยู่เสมอมา

กิจกรรมของมนุษย์เป็นปรากฏการณ์ที่กว้างขวางครอบคลุมอยู่ทุกชุมชน ไม่ว่า

จะเป็นในนครหลวง ในเมือง ในชนบท หรือในป่าตามเขา การศึกษาหาความรู้ในปัญหา และกิจกรรมเหล่านี้ย่อมจะต้องกระทำในขอบเขตที่กว้างขวางและอาศัยเทคนิคที่เชื่อถือได้ ผลที่ได้จึงจะเป็นที่ยอมรับนับถือในวงวิชาการ การวิจัยเชิงสำรวจเป็นระเบียบวิธีการอย่างหนึ่ง ที่ทำให้ได้มาซึ่งหลักฐานและข้อมูลต่าง ๆ ที่กว้างขวางและมีขอบเขตพอที่จะทำความเข้าใจและให้คำตอบต่อปัญหาปรากฏการณ์ในสังคมได้ จึงได้มีผู้นำวิธีการสำรวจมาใช้กันอย่างแพร่หลายในการศึกษาสถานะและความเป็นไปในทางสังคม ในระยะแรกมีการสำรวจชนิดสมบูรณ์แบบกล่าวคือ ทำการสำรวจตรวจสอบกันอยู่ทุกรูปทุกนามจนครบบริบูรณ์ อีกนัยหนึ่งคือ การแจกจ่ายทุก ๆ คนที่อยู่ในข่ายของการสำรวจ ต่อมาเมื่อวิชาสถิติศาสตร์เจริญมากขึ้น ประมาณ 50—60 ปีมานี้เอง การสำรวจจึงเริ่มแปรรูปจากการสำรวจทุกคนหรือทุกชั้นมาเป็นการสำรวจเฉพาะบางคนหรือบางส่วน ฉะนั้น การสำรวจแทนที่จะมุ่งเก็บข้อมูลของมวลทั้งหมดจึงแปรรูปมาเป็นการเลือกเก็บรวบรวมข้อมูลจากกลุ่มหนึ่งของมวลทั้งหมดหรือจากกลุ่มตัวอย่างที่เลือกมาจากมวล และเราเรียกวิธีนี้ว่า “การสุ่มตัวอย่าง”

วิธีการสุ่มตัวอย่างเป็นระเบียบวิธีการศึกษาอย่างหนึ่ง ที่ทำให้สามารถวางนัยโดยทั่วไป ในการอธิบายมวลปรากฏการณ์นั้นจากข้อมูลหรือข้อเท็จจริงที่สังเกตหรือรวบรวมได้จากส่วนหนึ่งของมวลปรากฏการณ์นั้น ฉะนั้น การศึกษาโดยวิธีการสุ่มตัวอย่างจึงมีหลักเกณฑ์และวิธีการที่แน่นอน การศึกษาแบบนี้รู้สึกว่าจะสะดวกและง่ายก็ เพราะเราใช้แต่เพียงข้อมูลส่วนหนึ่งเพื่อวิเคราะห์หาความถูกต้องและความเข้าใจในปรากฏการณ์ที่ศึกษา แต่ผลที่ได้รับก็เป็นที่น่าพอใจ ประสิทธิภาพของวิธีการเลือกศึกษาจากกลุ่มตัวอย่าง อาจพิจารณาได้จากเหตุการณ์ต่าง ๆ ในชีวิตประจำวัน เช่น เมื่อท่านป่วยและไปโรงพยาบาล นายแพทย์จะทำการตรวจโลหิต โดยใช้เข็มฉีดยาสุดโลหิตไปเพียง 5—10 ลบ. ซม. นายแพทย์ก็สามารถตรวจส่วนประกอบของเลือดในร่างกายทั้งหมดซึ่งมีประมาณ 4,000 ลบ. ซม. ได้ การเลือกชิมเงาะหรือลำไย 2—3 ผลจากเข่งหนึ่ง หรือการหยิบข้าวสารใส่ฝ่ามือตรวจดูก็พอจะตัดสินใจว่า ข้าวกระสอบนั้นควรจะจ่ายราคา 45 บาทต่อถังหรือไม่ วิธีการเหล่านี้เป็นมาตรการที่ปฏิบัติในชีวิตเศรษฐกิจและสังคมประจำวัน ตัวอย่างที่กล่าวเป็นการแสดงให้เห็นถึงการใช้ข้อมูลเพียงส่วนหนึ่ง เพื่ออธิบายคุณลักษณะของข้อมูลชุดนั้นทั้งชุด วิธีการสุ่มตัวอย่างเป็นวิธีการศึกษาค้นคว้าข้อมูลจากกลุ่มตัวอย่างเพื่อนำไปสู่การสรุปเกี่ยวกับลักษณะ

ส่วนรวมทั้งหมด ในทางสถิติเราเรียกส่วนหนึ่งของทั้งหมดว่า *กลุ่มตัวอย่าง* (sample) และส่วนรวมทั้งหมดเรียกว่า *ประชากร* (universe หรือ population) ค่าเฉพาะของข้อมูลที่ได้จากกลุ่มตัวอย่างเช่นค่าเฉลี่ย (mean) เรียกว่า *statistic* ค่าเฉพาะของข้อมูลที่ได้จากประชากรเรียกว่า *parameter* ฉะนั้น วัตถุประสงค์ของการสุ่มตัวอย่างก็คือการอุปนัย (inference) เกี่ยวกับค่า *parameter* (ซึ่งโดยปกติเป็นค่าที่ไม่มีใครทราบ) จากค่าที่คำนวณได้จากกลุ่มตัวอย่าง

ที่ได้กล่าวมาแล้วแสดงให้เห็นว่า การเลือกกลุ่มตัวอย่างเป็นขบวนการวิธีการที่ทำให้เราสามารถจะวางบทสรุปเกี่ยวกับคุณลักษณะบางประการของกลุ่มใหญ่ แม้ว่าเราจะเลือกศึกษาเฉพาะกลุ่มน้อย ๆ กลุ่มหนึ่งเท่านั้นก็ตาม ทั้งนี้คุณลักษณะจะเป็นการไม่น่าเชื่อในความถูกต้องอยู่บ้าง แต่เราก็ปฏิบัติเช่นนั้นอยู่ทุก ๆ วันโดยไม่เฉลียวใจแม้แต่เล็กน้อย ท่านจะพูดกับคนบางคนเพียงส่วนน้อย แล้วก็บอกว่าค่าครองชีพสูงขึ้นทุกวัน เรียนรู้อย่างเกี่ยวกับความเป็นไปของโลก คุยกับพวกพ่อค้า แม่ค้า กรรมกร คนขับรถบรรทุก ขับรถรับจ้าง ก็ทำให้ทราบว่าคนทั่วไปคิดอย่างไร รู้สึกอะไรและมีความเห็นอย่างไร เกี่ยวกับเรื่องราวและกระแสข่าวต่าง ๆ แต่อย่างไรก็ตามแม้การเรียนรู้จากคนกลุ่มน้อย เป็นสิ่งที่ง่ายและปฏิบัติได้ แต่การสุ่มตัวอย่างก็มีกฎเกณฑ์ที่ต้องระมัดระวังในการจะวางนัยโดยทั่วไปเกี่ยวกับเรื่องนั้นอยู่เหมือนกัน เป็นต้นว่า ถ้าต้องการทราบความคิดเห็นของคนในเรื่องหนึ่งเรื่องใดก็ต้องพยายามพูดกับคนในจำนวนที่มากพอๆ มิใช่พูดกับคน 2-3 คน ทั้งนี้ก็เพื่อจะทดสอบให้ครอบคลุมพิสัยอันกว้างขวางของความ คิดเห็น อันเป็น ลักษณะทั่วไปของกลุ่มคน หากการติดต่อสอบถามของเรามีแนวโน้มไปในทางที่จะสอบถามแต่คนที่ใกล้ชิดสนิทสนม ที่สนิทมีความคิดเห็นที่ได้ย่อมไม่น่าแสดงว่าเป็นความ คิดเห็นที่ถูกต้องของคนทั้งชุมชนนัก ถ้าเรามีโอกาสสอบถามผู้คนที่มีความคิดเห็นแตกต่างกันไปมากยิ่งขั้นก็ย่อมจะพิจารณาได้ว่า เราได้ใช้มาตรการที่น่าจะมีความถูกต้องมากกว่าเพื่อวัดความแตกต่างในความคิดเห็นของชุมชน การสำรวจวิจัยนั้นมุ่งหมายที่จะคาดคะเนคุณลักษณะที่ผิดแผกแตกต่างกันของประชากร ฉะนั้น จึงจำเป็นจะต้องทำการสุ่มตัวอย่างด้วยความระมัดระวังมากกว่าวิธีการสุ่มตัวอย่างในชีวิตประจำวัน ในทางปฏิบัตินั้นจำนวนและชนิดของประชากร

ในกลุ่มตัวอย่างมีความสำคัญต่อการวางนัยทั่วไปจากกลุ่มตัวอย่างทั่วๆ ไป กล่าวคือ จำนวนและชนิดของประชากรในกลุ่มตัวอย่างย่อมต้องมีเพียงพอที่จะเป็นตัวแทนของประชากร จึงจะทำให้การวางนัยโดยทั่วไปเกี่ยวกับประชากรนั้นมีความน่าเชื่อถือและวางอยู่บนหลักฐานที่มั่นคง

ก่อนที่จะได้พิจารณาต่อไปถึงทฤษฎีและหลักเกณฑ์เบื้องต้นของวิธีการสุ่มตัวอย่างเพื่อการวิจัยทางสังคมอันเป็นวัตถุประสงค์ของบทความนี้ ควรจะได้ทำความเข้าใจเกี่ยวกับคำเฉพาะบางคำที่ใช้เสียก่อน เพื่อที่จะได้เข้าใจทฤษฎีและหลักเกณฑ์บางประการได้ดียิ่งขึ้น

ประชากรและหน่วยประชากร

ในการสำรวจโดยวิธีการสุ่มตัวอย่างนั้น เรามีความประสงค์อย่างยิ่งที่จะได้กลุ่มตัวอย่างที่มีลักษณะเป็นตัวแทนของส่วนรวมทั้งหมด ส่วนทั้งหมดที่กล่าวถึงนี้ในทางวิจัยเรียกว่า "ประชากร" การให้คำนิยามคำว่า "ประชากร" นั้น บางทีไม่ใช่ของง่าย ๆ ตามที่คิดกันนัก ประชากรในความหมายตามปกติหมายถึง ประชาชนพลเมือง กลุ่มคนเป็นหมู่มีจำนวนมากๆ เรามักจะพูดกันว่าบ้านนี้มีคนอยู่ 5 คน แต่เราไม่พูดว่ามีประชากรหรือประชาชน 5 คน ถ้าเป็นหมู่บ้าน ตำบล จังหวัด หรือประเทศ เราจึงจะใช้คำว่า พลเมืองหรือประชาชนแล้วแต่กรณี ประชากรในความหมายทางวิชาสถิติศาสตร์ มีความหมายกว้างกว่าความหมายธรรมดา ประชากรในทางสถิติอาจหมายถึง คน สัตว์ หรือสิ่งของที่แสดงความเป็นกลุ่มก้อน หรือความอยู่ร่วมกันของแต่ละหน่วยที่มีลักษณะบางประการที่กำหนดไว้เหมือนกัน ฉะนั้น ประชากรในทางสถิติจึงไม่จำเป็นจะต้องหมายถึงบุคคลเสมอไป ในภาษาอังกฤษใช้คำว่า population หรือ universe เช่นเราจะศึกษาเรื่องราวของผู้อาศัยอยู่ในพระนคร ทุกคนที่อาศัยอยู่ในพระนครก็เป็นประชากร เราอาจกำหนดขอบเขตของประชากรว่าเป็นเด็กอายุต่ำกว่า 15 ซึ่งอาศัยอยู่ในอำเภอพระนคร หรือกำหนดขงรถยนต์ที่ผลิตออกจากโรงงานไฟร์สโตนในเดือนมิถุนายน หรือสุกรที่ส่งฆ่า ณ โรงฆ่าสัตว์ ของบริษัทสามัคคีฆ่าสุกรในเดือนมีนาคม หรือไม้สักที่ผลิตออกจากโรงงานไม้สักไทยในคาบระยะเวลาหนึ่ง สิ่งที่กำหนดขึ้นนั้นจะเป็นประชากรในความหมายทางสถิติทั้งสิ้น ฉะนั้น หากพิจารณาคุณะเห็นว่าประชากรเป็นสิ่งที่เราเป็นผู้กำหนดขึ้น เรากำหนดลักษณะของสิ่งที่เราจะต้องการศึกษา

เมื่อกำหนดคุณสมบัติหรือคุณลักษณะบางประการของสิ่งที่เราจะทำการสำรวจขึ้นแล้ว ทุกหน่วยของสิ่งนั้น ซึ่งมีคุณสมบัติดังกล่าวย่อรวมกันเป็นกลุ่มประชากร แต่ละหน่วยที่ประกอบกันขึ้นเป็นประชากรนั้น เรียกว่า หน่วยประชากร (element) หากจะพิจารณาทำให้ลึกซึ้งลงไปอีกเล็กน้อย ก็พอจะมองเห็นว่า หน่วยประชากร นั้นเป็นหน่วยที่เราต้องการจะได้ข้อมูลมาเพื่อทำการศึกษาความเป็นไปของประชากร กล่าวคือ เป็นหน่วยที่จะให้ข้อมูลเพื่อใช้ในการวิเคราะห์ เพราะฉะนั้น บางทีจึงเรียกหน่วยประชากรว่า หน่วยการวิเคราะห์ (Units of analysis) ลักษณะของหน่วยการวิเคราะห์นั้น ย่อมขึ้นอยู่กับวัตถุประสงค์ของการวิจัย

✕ การกำหนดคุณลักษณะหรือคุณสมบัติของหน่วยประชากรนั้น ย่อมมีความเกี่ยวข้องกับวัตถุประสงค์ของการสำรวจแต่ละครั้ง เช่น ต้องการสำรวจปัญหาในการปฏิบัติงานของกำนัน วัตถุประสงค์ที่จะสำรวจกำนัน ย่อมกำหนดหน่วยประชากรที่เราต้องการจะศึกษากล่าวคือ กำนัน ฉะนั้น กำนันซึ่งเป็นผู้ที่ได้รับการคัดเลือกจากหมู่บ้านในตำบลหนึ่ง และผ่านการแต่งตั้งจากกระทรวงมหาดไทยแล้ว จะเป็นหน่วยหนึ่งของประชากรที่ใช้ในการสำรวจ กำนันทุกคนทั่วประเทศย่อรวมกันเป็นกลุ่มประชากรที่เราต้องการ หากเราต้องการจะสำรวจว่า ร้านค้าย่อยที่กระทรวงเศรษฐกิจให้ความอนุเคราะห์ ปฏิบัติตามระเบียบข้อบังคับของกระทรวงหรือไม่ ร้านค้าย่อยที่ขึ้นทะเบียนอยู่ในความอุปการะของกระทรวงฯ ทุกร้านย่อรวมกันเป็นกลุ่มประชากรของการสำรวจ ตัวอย่างดังกล่าวชี้ให้เห็นถึงความสัมพันธ์ระหว่างจุดประสงค์ในการสำรวจกับการกำหนดหน่วยประชากร

เมื่อได้ประชากรตามลักษณะที่ต้องการแล้ว ปัญหาต่อไปก็คือ การกำหนดขอบเขตของประชากรที่ต้องการจะสำรวจ โดยทั่วไปนั้น ประชากรย่อมกระจุกกระจายกันอยู่ในพื้นที่ต่าง ๆ กันหรือไม่ก็เกิดขึ้นในช่วงเวลาที่แตกต่างกัน หากไม่กำหนดขอบเขตไว้ให้เป็นที่แน่นอน ย่อมจะทำให้การสำรวจเป็นไปด้วยความยากลำบาก ต้องใช้ผู้คนมาก เสียเวลา และค่าใช้จ่ายสูง มีหน้าซ้ำเมื่อได้ผลการสำรวจมาแล้วจะทำให้เกิดปัญหาว่าจะนำผลที่ได้ไปใช้กับใครที่ไหน การสำรวจย่อมจะไร้ความหมาย การกำหนดขอบเขตของประชากรมีความสำคัญต่อการสำรวจและเกี่ยวเนื่องกับปัจจัยหลายอย่าง เช่น วัตถุประสงค์ กำลังคน ค่าใช้จ่าย และเวลา ฉะนั้น การสำรวจแต่ละครั้งย่อมมีปัญหาเฉพาะเรื่อง ในการกำหนดขอบ

เขตของประชากรที่จะสำรวจ ผู้ทำการสำรวจจำเป็นต้องแก้ไขปัญหาเหล่านี้ให้ถูกต้อง และ
 เข้าใจปัญหานี้อย่างถ่องแท้เสียก่อน ทั้งนี้ ให้พิจารณาถึงความสามารถที่จะปฏิบัติได้หรือไม่
 เป็นประการสำคัญ

โดยปกติแล้วเราใช้ขอบเขตทางภูมิศาสตร์ เช่น ถนน แม่น้ำลำคลอง ภูเขา
 เขตการปกครองหรือระยะเวลาเป็นเครื่องกำหนดขอบเขตของการสำรวจ เช่น การสำรวจ
 กำหนด เราจะกำหนดขอบเขตเฉพาะกำหนดในภาคใดภาคหนึ่ง หรือทำการสำรวจร้านค้าย่อย
 เฉพาะในเขตอำเภอบางรัก สมมติว่าเราจะทำการสำรวจคนที่เรียนจบประถมศึกษาปีที่ 4 ไปแล้วมี
 ความสามารถในการอ่านและการเขียนหนังสืออย่างไร จากปัญหาดังกล่าวเราจะเห็นว่ากลุ่ม
 ประชากรมีอยู่มากมายทั่วประเทศ อันอาจทำให้การสำรวจเป็นไปได้ยาก หากจะทำการ
 สำรวจปัญหานี้จริงๆ เราอาจจะทำได้โดยกำหนดขอบเขตของประชากรใหม่ โดยใช้ขอบเขต
 ทางภูมิศาสตร์เป็นเครื่องกำหนด ซึ่งอาจเป็น อย่างเช่น กำหนดขอบเขตที่มีผู้คนไม่ค่อยหนา
 แน่นออกเสีย ตัดตำบลที่ห่างไกลออกไป กำหนดเขตสำรวจเฉพาะในอำเภอเมือง เป็นต้น
 นอกจากการใช้เส้นทางภูมิศาสตร์และเขตทางการปกครองเป็นเครื่องกำหนดขอบเขตของ
 ประชากรแล้ว เรายังมีวิธีการอื่นๆ อีก อาทิ การกำหนดช่วงอายุ รายได้ ขนาดของ
 ครอบครัว ช่วงเวลา เช่น ทำการสำรวจเฉพาะเด็กที่จบการศึกษาภาคบังคับแล้วอายุไม่
เกิน 20 ปี เด็กที่จบการศึกษาภาคบังคับแล้วไม่เกิน 5 ปี หรือหัวหน้าครอบครัวอายุไม่
 เกิน 30 ปี ฯลฯ อีกวิธีหนึ่งซึ่งอาจทำให้ขอบเขตของประชากรเล็กลงโดยการกำจัดบางส่วน
 ออกเสีย ซึ่งการทำเช่นนั้นจะไม่ทำให้ผลของการสำรวจเสียความสำคัญไป เช่น การสำรวจ
 สัมภาษณ์ประชากร พ.ศ. 2503 ขอบเขตการสำรวจไม่ได้รวมไปถึงชาวป่าเขาและชาวนา
 ฉะนั้น การสำรวจบางชนิดอาจจะไม่รวมประชากรบางจำพวก เช่น นักโทษ นักเรียน
 ทหาร ผู้ที่อาศัยอยู่ในหอพัก สถานที่ราชการ อายุต่ำกว่า 10 ปี อายุเกิน 60 ปี ผู้ที่อยู่
 ห่างไกลหรือเขตที่ประชาชนอยู่ไม่หนาแน่น ฯลฯ ทั้งนี้แล้วแต่กรณีและวัตถุประสงค์ของการ
 สำรวจและความสามารถในการปฏิบัติการสำรวจเป็นประเด็นสำคัญ

ข้อได้เปรียบของกลุ่มตัวอย่าง

เมื่อได้ทำการกำหนดลักษณะและขอบเขตของประชากรไว้เป็นที่แน่นอนแล้ว จะ

ทำให้ผู้สำรวจรู้แน่ชัดว่าหน่วยประชากรของการสำรวจ คือใครบ้าง เมื่อพิจารณาจุดปัจจัยต่าง ๆ ที่เกี่ยวข้องกับ การสำรวจจะทำให้ นักวิจัยตัดสินใจได้ว่า ควรจะทำการสำรวจแบบสำมะโน คือ การแจกแจงทุกหน่วยของประชากร หรือจะทำการสำรวจโดยใช้วิธีการสุ่มตัวอย่าง วิธีการสุ่มตัวอย่างนั้นมีข้อได้เปรียบอยู่หลายประการ เป็นที่เข้าใจกันว่าถ้ามวลประชากรมีขนาดใหญ่ ๆ และกระจายกันอยู่ในบริเวณกว้างแล้ว วิธีการสำรวจมักจะกระทำโดยการสุ่มตัวอย่างแทบทั้งสิ้น ข้อได้เปรียบของวิธีการสุ่มตัวอย่างนั้นมีดังนี้

1. วิธีการสุ่มตัวอย่าง เป็นวิธีการเก็บข้อมูลจากกลุ่มหนึ่งของประชากร ฉะนั้นจึงใช้เวลาและแรงงานน้อยกว่าการรวบรวมข้อมูลโดยวิธีการสำมะโน ซึ่งเก็บข้อมูลจากทุกหน่วยของประชากร ในการสำรวจบางอย่างหากใช้เวลามากไปในการเก็บและวิเคราะห์ข้อมูลย่อมจะทำให้ผลการสำรวจไม่ทันกาล กว่าผลการวิเคราะห์จะสำเร็จผลที่ได้อาจจะล้าสมัยไปแล้ว ในสังคมปัจจุบันนี้การประหยัดเวลาเป็นของสำคัญยิ่ง ภาวะของสังคมเคลื่อนไหวเปลี่ยนแปลงได้รวดเร็ว วิธีการสุ่มตัวอย่างมิใช่จะประหยัดเฉพาะเวลาการเก็บรวบรวมข้อมูลเท่านั้น หากประหยัดรวมไปถึงเวลาอธิกรณ์ การแยกประเภทและการจัดทำตารางต่าง ๆ ด้วย หรืออาจจะกล่าวได้ว่าเป็นการประหยัดเวลาในทุกขั้นตอนของขบวนการวิจัย เพราะกลุ่มตัวอย่างประกอบขึ้นด้วยหน่วยของประชากรน้อยกว่า

2. ถ้าได้ทำการวางแผนการสุ่มตัวอย่างที่ถูกต้อง การสำรวจโดยกลุ่มตัวอย่างจะสิ้นเปลืองค่าใช้จ่ายน้อยกว่าการทำสำมะโน กลุ่มตัวอย่างทำให้มีการสัมภาษณ์คนกลุ่มน้อยใช้เจ้าหน้าที่และพนักงานสำรวจและดำเนินการเกี่ยวกับข้อมูลน้อยกว่า ค่าใช้จ่ายจึงลดลงเป็นอย่างมาก ฉะนั้น วิธีการสุ่มตัวอย่างจึงทำให้โครงการวิจัยต่าง ๆ ที่มีเงินทุนอุดหนุนน้อยสามารถจะดำเนินการไปได้

3. ในบางกรณี การสำรวจโดยกลุ่มตัวอย่าง เป็นสิ่งที่ไม่สามารถจะหลีกเลี่ยงได้ เช่น ในกรณีที่เราจะต้องทำลายวัตถุที่เป็นหน่วยประชากรเพื่อตรวจสอบสมบัติบางประการที่ต้องการ การตรวจสอบในกรณีจึงต้องเลือกทำเฉพาะบางส่วนของประชากร หรือไม่ก็ทำลายประชากรทั้งหมดเสียเลย กลุ่มตัวอย่างแบบนี้อยู่ในพวกที่เรียกกันว่า destructive sampling เช่น หากจะสำรวจอายุของหลอดไฟฟ้า ความคงทนเป็นจำนวนกิโลเมตรของ

ขารรถยนต์ หากไม่ใช้การทดลองสำรวจกับกลุ่มตัวอย่าง นักวิจัยย่อมต้องทำลายผลิตภัณฑ์ทุกชิ้น

4. หากมีจำนวนคน ค่าใช้จ่าย และเวลาเท่ากันในการทำสำมะโนและการสำรวจโดยกลุ่มตัวอย่าง การประหยัดค่าใช้จ่ายโดยการใช้กลุ่มตัวอย่าง อาจจะนำส่วนที่ประหยัดไว้ไปใช้ในกรณีอื่น ๆ เช่น การทำการศึกษาเฉพาะกรณีเพื่อขยายขอบเขตของการวิจัยให้กว้างขวางลึกซึ้งยิ่งขึ้น ตระเตรียมการวิเคราะห์ข้อมูลให้ละเอียดละออกว่าที่คาดไว้แต่เดิม หรือนำไปใช้ในกรณีอื่นที่จะทำให้ผลการวิจัยมีคุณค่าและความสำคัญมากยิ่งขึ้น เช่น การเปรียบเทียบผลกับกลุ่มอื่นที่เปรียบเทียบกันได้ในสภาพการ เวลา และสถานที่

5. หากมองจากแง่บริหาร จะพบว่าในการสำมะโนนั้นจำเป็นต้องใช้เจ้าหน้าที่และผู้คนมากมาย ปัญหาเรื่องการฝึกอบรม การควบคุมการปฏิบัติงาน ปัญหาเรื่องเครื่องมือเครื่องใช้ ตลอดจนสถานที่ที่จะทำให้มองดูว่าทำสำมะโนเรื่องนี้ดูแทบจะเป็นไม่ได้ ความผิดพลาดอันเกิดจากการบริหารก็จะเกิดขึ้นเป็นเงาตามตัว อันจะทำให้ผลจากการสำมะโนไม่มีประสิทธิภาพเท่าที่ควร ปัญหาเหล่านี้จะทำให้เกิดความจำเป็นและมองเห็นประโยชน์ของวิธีการสุ่มตัวอย่าง

6. การใช้กลุ่มตัวอย่าง ทำให้มีเวลาพอที่จะตรวจสอบและแก้ไขข้อผิดพลาดต่าง ๆ ในขบวนการวิจัย การเอาใจใส่ในขั้นตอนต่าง ๆ อย่างเพียงพอ ย่อมเป็นหลักประกันที่มั่นใจอันหนึ่งในความถูกต้องของผลการวิจัย ตารางการแจกแจงข้อมูลและการวิเคราะห์ข้อมูลย่อมเป็นที่เชื่อถือได้มากกว่าข้อมูลต่างๆที่ได้รวบรวมมาด้วยความระมัดระวังที่น้อยกว่า

ข้อได้เปรียบของวิธีการสุ่มตัวอย่างตามที่กล่าวมาแล้วพอสรุปได้ว่า วิธีการสุ่มตัวอย่างทำให้ประหยัดเวลา ค่าใช้จ่าย และแรงงานได้เป็นอย่างมาก การทำสำมะโนแม้ว่าผลที่ได้จะถูกต้องแน่นอนกว่า แต่การทำสำมะโนมีโอกาสเกิดความผิดพลาดได้มากทั้งในด้านการบริหารและในด้านเทคนิค หากคิดว่าผู้นำเวลา เงินทอง และแรงงานที่จะเสียไปในการทำสำมะโนไปใช้จ่ายเพิ่มในวิธีการสุ่มตัวอย่าง สิ่งเหล่านี้ย่อมนำไปสู่การเพิ่มความถูกต้องความเชื่อถือ และความสมบูรณ์ของผลที่ได้จากกลุ่มตัวอย่างได้เป็นอย่างดี

อย่างไรก็ตาม วิธีการสุ่มตัวอย่างไม่ใช่จะไม่มีข้อจำกัด หรือเป็นของที่จัดทำขึ้นได้ง่าย ๆ ตามที่คิดกัน กลุ่มตัวอย่างที่จะเป็นตัวแทนของประชากรได้ ต้องได้จากการพิจารณาและการวางแผนแบบในการสุ่มตัวอย่างที่ถูกต้อง และจะต้องดำเนินการไปตามแบบแผนที่วางไว้ มิฉะนั้นผลที่ได้ย่อมไม่ถูกต้อง และนำไปสู่การเข้าใจผิดต่าง ๆ ในบางกรณีวิธีการสุ่มตัวอย่างไม่สามารถจะคลุมคุณสมบัติบางประการที่หาได้ยากในกลุ่มประชากร เช่น การคาดคะเนอายุ 90 ปี ในกลุ่มประชากร 5000 ย่อมทำได้โดยยากด้วยกลุ่มตัวอย่าง นอกจากนี้หากการสำรวจมีความประสงค์ต้องการแยกแยะแจกแจงข้อมูลให้ละเอียดถี่ถ้วน การใช้วิธีสุ่มตัวอย่างย่อมจะให้ผลไม่เป็นที่น่าเชื่อถือ เพราะการแจกแจงข้อมูลอย่างละเอียดจะทำให้จำนวนประชากรในกลุ่มตัวอย่างถูกย่อยลงไปในระดับต่าง ๆ ซึ่งแต่ละชั้นจะมีจำนวนไม่พอเพียงที่จะทำให้การวิเคราะห์ทางสถิติเป็นที่น่าเชื่อถือได้ นอกเสียจากจะใช้กลุ่มตัวอย่างขนาดใหญ่มาก ซึ่งก็ดูเป็นการทำสำมะโนไป นอกจากข้อจำกัดต่าง ๆ ดังกล่าวมาแล้ว ยังปรากฏว่าการสำรวจโดยการใช้กลุ่มตัวอย่างเป็นบริการของผู้เชี่ยวชาญ มีผู้กล่าวว่าในปัจจุบันนี้เรายังขาดผู้เชี่ยวชาญทางการสุ่มตัวอย่างเป็นอย่างมาก จริงอยู่เรามีนักสถิติอยู่บ้าง แต่การสุ่มตัวอย่าง โดยเฉพาะการสุ่มมวลมนุษยย์ต้องการความชำนาญเป็นพิเศษเกี่ยวกับข้อมูลทางสังคมและความรอบรู้เกี่ยวกับความสัมพันธ์เกี่ยวเนื่องกันระหว่างข้อมูลทางสังคมเหล่านั้น

ในการสุ่มตัวอย่างนั้นผู้ทำการสำรวจพึงระลึกว่ากลุ่มตัวอย่างที่กินนั้น ย่อมจะได้รับเลือกมาจากการที่ได้ศึกษาพิจารณาลักษณะของประชากรที่จะทำการสำรวจแล้วเท่านั้น ในบางครั้งจำเป็นจะต้องทำการสำรวจล่วงหน้าเป็นโครงการเล็ก ๆ เพื่อตรวจหาคุณลักษณะทั่วๆ ไปบางประการเสียก่อนว่า ประชากรมีความผิดแผกแตกต่างกันมากน้อยเพียงไร สำรวจดูว่ามีอะไรบ้างที่จะเป็นประโยชน์แก่อุปการะในการทำแผนแบบการสุ่ม เช่น บัญชีรายชื่อประชากร แผนที่ แผนแบบที่คนอื่นเคยทำไว้แล้ว และสำรวจตรวจหาประสพการณ์จากผู้ที่เคยทำโครงการวิจัยอื่น ๆ มาแล้ว วิธีการเหล่านี้จะเป็นเครื่องประกันให้ผู้สำรวจได้เลือกกลุ่มตัวอย่างที่ดี ซึ่งเป็นตัวแทนของกลุ่มประชากรได้เป็นอย่างดี

ลักษณะของกลุ่มตัวอย่างที่ดี

ในการสำรวจโดยวิธีสุ่มตัวอย่างนั้น ผู้ทำการสำรวจจำเป็นต้องใช้ขบวนการต่าง ๆ เพื่อเลือกหาตัวอย่างที่ดี ซึ่งเป็นตัวแทนของประชากร ความเชื่อถือได้ของกลุ่มตัวอย่างทำให้เราสามารถจะวางนัยโดยทั่วไปเกี่ยวกับประชากรที่เรามุ่งจะศึกษาให้เป็นที่ยอมรับนับถือได้ ความเชื่อถือได้ของกลุ่มตัวอย่างดังกล่าว มิได้ขึ้นอยู่กับวิธีการเลือกสรรหาหน่วยประชากรที่จะให้ข้อมูลที่ถูกต้องเที่ยงตรง กลุ่มตัวอย่างที่ดีขึ้นอยู่กับวิธีการเลือกแต่ละหน่วยของประชากรขึ้นเป็นกลุ่มตัวอย่าง เพื่อให้ได้กลุ่มตัวอย่างที่ดี วิธีการเลือกกลุ่มตัวอย่างย่อมจะต้องวางรากฐานอยู่บนทฤษฎีบางประการ กล่าวคือ

1. สามารถอำนวยความสะดวกถึงขนาดของกลุ่มตัวอย่างที่ต้องการ
2. สามารถตรวจหา โฉลก (chance) หรือความน่าจะเป็นไปได้ของแต่ละหน่วยประชากรที่จะถูกเลือกเป็นกลุ่มตัวอย่าง
3. สามารถตรวจสอบความผิดพลาดอันเป็นผลสืบเนื่องมาจากการเลือกกลุ่มตัวอย่างแทนการที่จะใช้ประชากรทั้งหมด
4. สามารถจะพิจารณาระดับแห่งความมั่นใจในการวางนัยโดยทั่วไป (generalization) เกี่ยวกับประชากรจากกลุ่มตัวอย่าง

มาตรการที่วางไว้เป็นหลักเกณฑ์ในการเลือกกลุ่มตัวอย่างนี้เป็นหลักเกณฑ์ที่ทำให้เกิดทฤษฎีการสุ่มตัวอย่าง กลุ่มตัวอย่างที่ได้จากการเลือกตามกฎเกณฑ์ดังกล่าวนี้เมื่อเรียกกันโดยทั่วไปว่า กลุ่มตัวอย่างที่มีความน่าจะเป็นไปได้ (probability sample) วิธีการที่ดำเนินไปเพื่อให้ได้กลุ่มตัวอย่างเราเรียกกันว่า วิธีการสุ่มตัวอย่าง จากกฎเกณฑ์ดังกล่าวจะเห็นว่า การสุ่มในที่นี้มีใช้หมายถึงการสุ่มหรือการเตาสุ่มในความหมายธรรมดา ซึ่งมีนัยไปในทำนองของความบังเอิญปราศจากหลักเกณฑ์ที่จะยึดถือ แล้วแต่โชคชะตาจะพาไปตามยถากรรม หากแต่ว่าการสุ่มในที่นี้เป็น การสุ่มที่เป็นไปตามกฎเกณฑ์หรือทฤษฎีที่ได้วางไว้เป็นบรรทัดฐาน กลุ่มตัวอย่างที่ได้จากทฤษฎีดังกล่าวนี้เท่านั้นที่ทำให้เราสามารถคาดคะเนความผิดพลาดที่เกี่ยวข้องกับการเลือกกลุ่มตัวอย่างได้ การจะให้ได้กลุ่มตัวอย่างที่ดีอาจจะเสียค่าใช้จ่ายมากพอสมควร เมื่อพิจารณาใช้วิธีการสุ่มตัวอย่างจึงจำเป็นจะต้องคำนึงถึง

ความเปลี่ยนแปลงต่าง ๆ ไว้ด้วย โดยทั่วไปแล้วการวางแผนแบบ (design) เพื่อเลือกกลุ่มตัวอย่าง จึงน่าจะได้พิจารณาถึงความต้องการและความถูกต้องทางทฤษฎีและข้อจำกัดในเชิงปฏิบัติด้วย เมื่อพิจารณาในแง่ปฏิบัติ กลุ่มตัวอย่างที่ดีคือกลุ่มตัวอย่างซึ่งสามารถจะให้ข้อมูลที่ต้องการภายในขอบเขตของความผิดพลาดที่สามารถจะรับได้ตามที่คาดหมายไว้แล้วด้วยราคาค่าใช้จ่ายที่ต่ำที่สุด หรืออีกนัยหนึ่งอาจจะกล่าวได้ว่า กลุ่มตัวอย่างที่ดี คือกลุ่มตัวอย่างซึ่งภายในวงเงินที่กำหนด สามารถจะให้ข้อมูลที่ต้องการได้โดยมีความผิดพลาดอันสืบเนื่องมาจากกลุ่มตัวอย่าง (sampling error) น้อยที่สุด

เป้าหมายของกลุ่มตัวอย่าง

การสำรวจโดยวิธีการสุ่มตัวอย่างนั้น มีจุดประสงค์สำคัญอยู่ ๒ ประการ คือ การคาดคะเนหรือคาดหมายคุณลักษณะบางประการหรือค่าของประชากร (population estimate) และการทดสอบสมมติฐานบางประการเกี่ยวกับประชากรที่ต้องการสำรวจ (testing hypothesis)

1) การคาดคะเนค่าของประชากร ตามที่ได้กล่าวมาแล้วในการสำรวจศึกษาหาความเป็นไปและคุณลักษณะบางประการของประชากรนั้น อาจทำได้ ๒ แบบ คือวิธีการสำมะโน และวิธีการสุ่มตัวอย่าง การสุ่มตัวอย่างนั้นเป็นการศึกษาเพียงส่วนหนึ่งของประชากรที่เรากำหนดขอบเขตไว้ โดยมีจุดประสงค์ที่จะใช้ผลที่ได้จากกลุ่มตัวอย่างวางนัยโดยทั่วไป (generalization) เกี่ยวกับมวลประชากรหรือใช้ผลที่ได้จากกลุ่มตัวอย่างคาดคะเนใกล้เคียงค่าของประชากร ตามปกติในทางสถิติใช้ค่าเฉลี่ย อัตราส่วนหรือร้อยละที่คำนวณให้ได้จากกลุ่มตัวอย่างเป็นเครื่องมือในการประเมินค่าของประชากร (parameter) ตามธรรมชาติค่าของประชากรเหล่านี้เป็นค่าที่ไม่มีใครทราบ เว้นเสียแต่จะทำสำมะโนแจกนับประชากรนั้นเป็นรายตัว การคาดคะเนใด ๆ นั้นย่อมมีความผิดพลาดเกิดขึ้นไม่มากก็น้อย การคาดคะเนของประชากรจากค่าของกลุ่มตัวอย่างก็เช่นกัน ย่อมจะมีการผิดพลาดเกิดขึ้นจากทฤษฎีการสุ่มตัวอย่าง ความผิดพลาดของกลุ่มตัวอย่างมีอยู่ ๒ ประการ คือความผิดพลาดอันสืบเนื่องมาจากการสุ่มตัวอย่าง (sampling error) และความผิดอย่างอื่น ๆ (non sampling

error) ความผิดพลาดอันเนื่องมาจากกลุ่มตัวอย่างนั้น วิชาสถิติอาจจะทดสอบได้ โดยใช้มาตรการที่เรียกว่า การวัดความผิดพลาดมาตรฐาน (standard error) จากความผิดพลาดมาตรฐานนี้เอง ทฤษฎีการสุ่มตัวอย่างสามารถอธิบายถึงความแม่นยำ (precision) ของการคาดคะเนค่าของประชากรได้

2. การทดสอบสมมติฐาน การสำรวจวิจัยในบางกรณี เราอาจต้องการพิจารณาและศึกษาว่าความคิดเห็นทั่ว ๆ ไปเกี่ยวกับประชากรในเรื่องหนึ่งเรื่องใดนั้น จะถูกต้องเพียงไรหรือไม่เมื่อภาวะทางสังคมและเศรษฐกิจเปลี่ยนแปลงไป กระบวนแห่งความเป็นไปของสังคมจะยังคงรูปอยู่ตามที่คิดไว้แต่เดิมหรือจะเป็นไปตามที่คาดคะเนขึ้นใหม่หรือไม่เพียงใด ฉะนั้น จำเป็นจะต้องทำการสำรวจเพื่อทดสอบข้อสมมติเหล่านี้ เช่น เราตั้งสมมติฐานว่า ครอบครัวในเขตเทศบาลนครกรุงเทพ ฯ อย่างน้อย 50% มีเครื่องรับโทรทัศน์ หากเราสำรวจโดยใช้กลุ่มตัวอย่าง เลือกครอบครัวในเขตเทศบาลมากลุ่มหนึ่ง คำนวณหาครัวเรือนที่มีเครื่องรับโทรทัศน์ ผลของการสำรวจแสดงว่า 38% ของครัวเรือนในกลุ่มตัวอย่างมีเครื่องรับโทรทัศน์ ปัญหาต่อไปก็คือการที่จะวิเคราะห์ดูว่า ผลที่ได้จากการสำรวจนั้นเป็นข้อสนับสนุนหรือขัดแย้งสมมติฐานที่ตั้งไว้ว่า 50% ของครัวเรือนในเขตเทศบาลนครกรุงเทพ ฯ มีเครื่องรับโทรทัศน์ ในการวิเคราะห์ปัญหาดังกล่าวต้องอาศัยหลักเกณฑ์และทฤษฎีบางประการที่จะพิจารณาตัดสินว่าผลลัพธ์จากกลุ่มตัวอย่างมีความผิดพลาดเบี่ยงเบนไปจากค่าที่สมมติขึ้นมีความสำคัญเพียงไร หลักเกณฑ์ที่ใช้ก็คือ มาตรการวัดความแม่นยำที่กล่าวในวรรคที่แล้ว

การคาดคะเนค่าของประชากร และการทดสอบสมมติฐานดังกล่าวได้พัฒนาไปสู่ส่วนสำคัญ ๆ ของทฤษฎีทางสถิติศาสตร์ จุดประสงค์ทั้งสองนี้มีความสัมพันธ์เกี่ยวเนื่องกันอย่างใกล้ชิด กลจักรที่จะนำไปสู่ความสัมฤทธิ์ผลของจุดประสงค์ทั้งสองย่อมขึ้นอยู่กับมาตรการวัดความผิดพลาดมาตรฐาน

เนื่องจากการคาดคะเนค่าของประชากร และการทดสอบสมมติฐานเป็นแนวความคิดที่ต้องการทดสอบความถูกต้องแม่นยำของผลที่ได้ ดังนั้นข้อแตกต่างต่าง ๆ ที่เกี่ยว

กับประชากรอันสืบเนื่องมาจากผลลัพธ์ของกลุ่มตัวอย่างจึงเป็นข้อแถลงในเชิงความน่าจะเป็น
 ไปได้ของผลมากกว่าจะเป็นการแถลงถึงความจริง พึงระลึกว่าในการสุ่มตัวอย่างนั้นเรา
 พิจารณาเลือกศึกษาเพียงส่วนหนึ่งของประชากรเท่านั้น ฉะนั้น เราได้ละทิ้งภาคพรรณนา
 เกี่ยวกับมวลประชากรและความแน่นอนของข้อมูลอันเกี่ยวกับประชากรไว้เบื้องหลัง การ
 สืบตรวจโดยใช้กลุ่มตัวอย่างจึงเป็นวิธีการเลือกสรรเพื่อทำการอุปนัย (inference) หรือสรุป
 เกี่ยวกับสภาวะความน่าจะเป็นได้ของเรื่องที่ศึกษา สมมติว่าการสำรวจข้าราชการกระทรวง
 หนึ่งซึ่งมีจำนวน 5,000 คน การสำรวจเป็นแบบสำมะโน คือสัมภาษณ์ข้าราชการทุกคน
 และพบว่า 2,500 คน สูบบุหรี่ เราย่อมกล่าวได้ในเชิงที่เป็นความจริงว่าข้าราชการกระทรวง
 นี้สูบบุหรี่ 50% แต่ถ้าเราสัมภาษณ์ข้าราชการเพียง 4,998 คน และพบว่า 2,499 คน สูบบุหรี่
 กล่าวคือ 50% เหมือนกัน ข้อแถลงในกรณีหลังนี้ เรากล่าวได้ในรูปของการคาดคะเน
 (estimate) เท่านั้น หากใช้ข้อแถลงที่เป็นความจริงไม่ ทั้งนี้เพราะผลที่ได้จากอันหลังยังขาด
 ความรู้เกี่ยวกับข้อมูลของประชากรบางหน่วย ฉะนั้น การวางนัยทั่วไป (generalization)
 เกี่ยวกับประชากร โดยใช้กลุ่มตัวอย่างเป็นปทัสถานจึงจำเป็นต้องแถลงในเชิงความน่าจะเป็น
 เป็นไปได้

อีกประการหนึ่งขอให้พึงระลึกว่า หลักการต่าง ๆ ที่ใช้กับกลุ่มตัวอย่างที่กล่าวมา
 แล้วนั้นขึ้นอยู่กับข้อสมมติที่ว่า กลุ่มตัวอย่างที่ได้นั้นมาจากวิธีการสุ่มตัวอย่างดังกล่าวแล้ว
 ในตอนต้นเท่านั้น มิใช่จะใช้ได้กับกลุ่มตัวอย่างที่คัดเลือกมาโดยวิธีการอย่างอื่น

ความถูกต้อง (Accuracy) ความเอนเอียง (Bias) และความแม่นยำ (Precision)

คนธรรมดาส่วนมากมักจะมีความสงสัยและไม่ค่อยเชื่อว่า ทำไมการไปเลือก
 สัมภาษณ์หรือเก็บข้อมูลมาจากส่วนน้อยหรือบางส่วนจึงจะทำให้เราสามารถอธิบายเรื่องของ
 กลุ่มประชากรทั้งหมดได้ สำหรับพวกที่มีความเชื่ออยู่ว่า วิธีการกลุ่มตัวอย่างทำให้เรา
 สามารถวางนัยทั่วไปเกี่ยวกับประชากรที่เราศึกษาได้ แต่ก็สงสัยว่า เราจะเลือกกลุ่มตัวอย่าง
 ที่เป็นตัวแทนของมวลประชากรได้อย่างไร เพราะพิจารณาดูแล้วมันก็ออกจะเหลือเชื่อที่จะ

เลือกกลุ่มตัวอย่างมา 500 เพื่อจะใช้แทนประชากร 50,000 คน หรือกลุ่มตัวอย่าง 5,000 คน แทนประชากรทั้งประเทศ

ความรู้เกี่ยวกับวิธีการสุ่มตัวอย่างนั้น อาจจะแสดงให้เห็นได้ชัดและเข้าใจได้ดียิ่งขึ้น หากได้ลองพิจารณาจากกลุ่มประชากรขนาดเล็ก ๆ เป็นแบบตัวอย่าง ทำการเลือกกลุ่มตัวอย่างตามแบบที่เรียกกันว่า วิธีการสุ่มตัวอย่างแบบง่าย (simple random sampling) กล่าวคือ ทำการเลือกกลุ่มตัวอย่างโดยให้แต่ละกลุ่มตัวอย่างที่อาจจะเป็นไปได้มีโอกาสจะถูกเลือกเท่ากัน ขอให้พิจารณาดังตัวอย่างสมมติต่อไปนี้

สมมติว่ามีมวลประชากรขนาดเล็ก ประกอบด้วย A, B, C และ D รวม 4 คน ซึ่งมีอายุ 15, 17, 18 และ 22 ตามลำดับ อายุโดยเฉลี่ยของประชากรกลุ่มนี้จะเท่ากับ $15+17+18+22 \div 4 = 18$ ดังนั้น ค่าประชากรที่เราจะทำการคาดคะเนเท่ากับ 18 ขอให้สังเกตว่าในทางปฏิบัตินั้น ค่าของประชากรดังกล่าวนี้เราไม่ทราบ แล้วก็อาจจะไม่มีใครทราบด้วย นอกจากเราจะทำสำมะโนแจงนับประชากร ฉะนั้น เราจึงจำต้องหาทางคาดคะเนค่าของประชากรโดยใช้ผลที่ได้จากกลุ่มตัวอย่าง ส่วนการคาดคะเนนั้นจะมีความถูกต้อง มีความลำเอียง หรือความแม่นยำแค่ไหน เพียงใดนั้นจะได้นำมาพิจารณากันต่อไป ฉะนั้น ในขั้นนี้จึงจำเป็นสมมติแบบประชากรขึ้นและรู้ค่าของประชากร คือ อายุเฉลี่ยเท่ากับ 18

สมมติต่อไปว่า เราต้องการจะเลือกกลุ่มตัวอย่างซึ่งมีขนาดเพียง 2 คือกลุ่มตัวอย่างที่ประกอบด้วยหน่วยประชากร 2 หน่วย จากกลุ่มประชากรเล็ก ๆ ซึ่งมี 4 หน่วยดังกล่าวมาแล้ว เพื่อคำนวณคาดคะเนค่าของประชากร ปัญหาต่อไปอยู่ที่ว่า เราจะเลือกกลุ่มตัวอย่าง 2 หน่วยมาได้อย่างไร กลุ่มตัวอย่างที่จะถูกเลือกมานั้นจะประกอบด้วยหน่วยใดบ้างจากประชากรทั้งหมด จะเป็น A กับ B, B กับ C หรือ C กับ D หรือจะเป็นอย่างอื่น วิธีการสุ่มตัวอย่างแบบง่ายหรือ Simple Random Sampling นั้น เป็นวิธีการเลือกที่จัดให้กลุ่มตัวอย่างที่อาจจะเป็นไปได้ทุกกลุ่มตัวอย่างมีโอกาสที่จะถูกเลือกเท่ากัน ในขั้นนี้เรื่องที่น่าพิจารณาคือ กลุ่มตัวอย่างที่อาจจะเป็นไปได้ อันหมายถึงการที่แต่ละหน่วยของประชากร

สามารถจะผันแปรรวมกันเป็นกลุ่มตัวอย่างขึ้นมา เท่าที่สามารถจะผันแปรไปได้ในขนาดของกลุ่มตัวอย่างที่กำหนดไว้ สำหรับกรณีนี้ กลุ่มตัวอย่างอาจจะเป็น AB, AC, AD, BC, BD, หรือ CD กล่าวคือ จากประชากร 4 หน่วยจะมีกลุ่มตัวอย่างที่มีขนาด 2 หน่วยประชากร และที่อาจจะเป็นไปได้ 6 กลุ่มตัวอย่างด้วยกัน ฉะนั้น เมื่อทำการสุ่มตัวอย่างเราอาจจะสุ่มกลุ่มใดกลุ่มหนึ่งใน 6 กลุ่มนั้นขึ้นมาก็ได้

หากเราสุ่มได้ AC ขึ้นมาเป็นกลุ่มตัวอย่าง เมื่อทำการศึกษาข้อมูลจาก A และ C พบว่า A และ C มีอายุ 15 และ 18 ตามลำดับ ค่าเฉลี่ยอายุจากกลุ่มตัวอย่างจะเท่ากับ $\frac{15 + 18}{2} = 16.5$ ตามวิธีการสุ่มตัวอย่าง เราจะใช้ค่า 16.5 เป็นค่าสำหรับการคาดคะเนค่าของประชากร แต่จากตัวอย่างที่เราสมมติขึ้นนั้นเราทราบแล้วว่า ค่าอายุเฉลี่ยของประชากรเป็น 18 ฉะนั้น จะเห็นว่าผลลัพธ์จากกลุ่มตัวอย่างใช้คาดคะเนของประชากรไม่ถูกต้องนัก หากกลุ่มตัวอย่างที่เราเลือกได้เกิดเป็น C กับ D ซึ่งมีอายุ 18 และ 22 ค่าเฉลี่ยก็จะเท่ากับ 20 หากใช้ค่าของกลุ่มตัวอย่างนี้คาดคะเนค่าของประชากร จะเห็นว่าความถูกต้องยิ่งน้อยลงไปกว่าค่าของกลุ่มแรกอีก แต่ถ้ากลุ่มตัวอย่างที่สุ่มได้เกิดเป็น B กับ C ซึ่งมีอายุ 17 และ 18 อายุเฉลี่ยเป็น 17.5 หากใช้ค่าที่ได้นี้คาดคะเนค่าของประชากรจะเป็นค่าที่ใกล้เคียงมาก ฉะนั้น จะเห็นได้ว่าการถูกต้อง (accuracy) ของการคาดคะเนหมายถึง การคาดคะเนที่มีความใกล้เคียงค่าที่แท้จริงของประชากร โดยทั่วไปนั้นเราไม่ทราบค่าที่ถูกต้องแท้จริงของประชากรนั้น ฉะนั้น เราจึงไม่สามารถประเมินการคาดคะเนจากกลุ่มตัวอย่างได้ว่าถูกต้องเพียงไร ความถูกต้องที่กล่าวถึงนี้ จึงเป็นความถูกต้องที่น่าจะเป็นไปได้ (probable accuracy) ซึ่งในทางสถิติศาสตร์ เรียกว่า ความแม่นยำ (precision)

เมื่อเราพิจารณาค่าต่อไปจะพบว่า ตามวิธีการเลือกกลุ่มตัวอย่างแล้ว เราจะได้ค่าจากกลุ่มตัวอย่าง (statistic) เพื่อใช้คาดคะเนอยู่ 6 ค่าด้วยกัน สุดแต่โอกาสหรือโผลงที่เราจะเลือกได้กลุ่มตัวอย่างใดขึ้นมา เมื่อพิจารณาทุกกลุ่มตัวอย่างทั้ง 6 กลุ่มจะพบว่าทั้ง 6 กลุ่มให้ค่าแตกต่างกันทั้งสิ้น

กลุ่มตัวอย่างที่ 1	AB	อายุ 15 และ 17	อายุเฉลี่ย 16
„ 2	AC	„ 15 „ 18	„ 16.5
„ 3	AD	„ 15 „ 22	„ 18.5
„ 4	BC	„ 17 „ 18	„ 17.5
„ 5	BD	„ 17 „ 22	„ 19.5
„ 6	CD	„ 18 „ 22	„ 20
รวม			108.0

หากใช้มโนภาพทำการสุ่มตัวอย่างขึ้นมาทีละคู่ จดจำนวนครั้งของแต่ละคู่ไว้ แล้วใส่คู่ที่สุ่มมาได้ลงไปทีเดิม ทำการสุ่มอีกและทำเช่นนี้เรื่อย ๆ ไปจนไม่มีที่สิ้นสุด เราจะพบว่าเมื่อกลุ่มตัวอย่างที่อาจจะเป็นไปได้ทุกกลุ่มมีโอกาสที่จะถูกเลือกมาเป็นตัวอย่างเท่ากันแล้ว ในที่สุดกลุ่มตัวอย่างแต่ละกลุ่มจะถูกเลือกขึ้นมาเป็นจำนวนครั้งเท่า ๆ กัน เพราะตามทฤษฎีความน่าจะเป็น (probability theory) แต่ละกลุ่มตัวอย่างดังกล่าวมีโอกาสจะถูกเลือกเท่ากันคือ $1/6$ ฉะนั้น ผลเฉลี่ยของค่าที่จะใช้คาดคะเนประชากรที่คำนวณได้จากกลุ่มตัวอย่างที่อาจจะเป็นไปได้ทั้งหมด อายุเฉลี่ย $108 \div 6 = 18$ ซึ่งค่าเฉลี่ยดังกล่าวนี้เท่ากับค่าเฉลี่ยของประชากร คือ $\frac{15+17+19+22}{4} = 18$ ค่าเฉลี่ยที่ได้จากกลุ่มตัวอย่างที่มีจำนวนไม่สิ้นสุดตามที่ยกตัวอย่างมานี้ เรียกว่า ค่าที่คาดหวังเพื่อใช้ในการคาดคะเน (The expected value of the estimate) หากค่าที่คาดหวังนี้เท่ากับค่าของประชากรจริง ๆ ในกรณีเช่นนี้เราเรียกว่า วิธีการสุ่มตัวอย่างไม่มีอคติ หรือไม่มีอคติในการเลือกกลุ่มตัวอย่าง ฉะนั้น ผลแตกต่างระหว่างค่าที่คาดหวังและค่าแท้จริงของประชากรจึงหมายถึงความเบี่ยงเบนหรืออคติ การสุ่มตัวอย่างที่มีอคติหรือทำให้เกิดกลุ่มตัวอย่างเบี่ยงเบนนั้น จะพบในกรณีที่หน่วยหนึ่งหน่วยใดของประชากร เช่น คนที่มีอายุ 15 ปี ไม่มีโอกาสจะถูกเลือกมาเป็นกลุ่มตัวอย่าง^๕ อาจจะเป็นเพราะเหตุใด ๆ ก็ตาม ในกรณีเช่นนี้ กลุ่มตัวอย่างที่อาจจะถูกเลือกจะมีเพียง 3 กลุ่ม คือ BC, BD และ CD ซึ่งมีค่าอายุเฉลี่ยเท่ากับ 17.5, 19.5 และ 20 ตามลำดับ ฉะนั้น ค่าที่คาดหวัง (The expected value) จึงเท่ากับ $\frac{17.5+19.5+20}{3} = 19$ ซึ่งสูงกว่าค่าเฉลี่ยของประชากร ในกรณีนี้ความเบี่ยงเบนจึงเท่ากับ 1

จากตัวอย่างที่สมมติขึ้นดังกล่าวมาแล้ว จะพบว่าค่าอายุเฉลี่ยของประชากรเท่ากับ 18 แต่กลุ่มตัวอย่างซึ่งมีขนาด 2 หน่วยนั้น ไม่มีกลุ่มตัวอย่างใดใน 6 กลุ่มนั้น ให้ค่าเฉลี่ยเท่ากับ 18 เลยสักกลุ่มเดียว กล่าวคือ ถ้าสุ่มกลุ่มหนึ่งกลุ่มใดขึ้นมาหาค่าเฉลี่ยแล้วใช้ค่าเฉลี่ยนั้นคาดคะเนค่าประชากร การคาดคะเนจะผิดพลาดทั้งนั้น ผิดมากบ้างน้อยบ้างแล้วแต่จะสุ่มได้กลุ่มใดขึ้นมา ทั้ง ๆ ที่วิธีการสุ่มตัวอย่างก็ไม่มีอคติ ดังนั้น ตัวอย่างดังกล่าวแสดงให้เห็นหลักการที่สำคัญประการหนึ่งในทฤษฎีการสุ่มตัวอย่าง คือ ค่าเฉลี่ยของค่าคาดคะเนที่ได้จากกลุ่มตัวอย่างทั้งหมด จะมีค่าเท่ากับค่าเฉลี่ยของประชากร ฉะนั้น จึงเป็นการแสดงให้เห็นว่า กลุ่มตัวอย่างกลุ่มใดก็ตามอาจจะให้ค่าที่ใช้คาดคะเนไม่ถูกต้องทั้ง ๆ ที่วิธีการสุ่มตัวอย่างไม่มีอคติ จากตัวอย่างที่สมมติขึ้น จะเห็นได้ว่าการคาดคะเนค่าของประชากรโดยใช้ผลลัพธ์จากกลุ่มตัวอย่างที่ 1 นั้น เป็นการคาดคะเนที่ไม่ถูกต้อง เพราะอายุเฉลี่ยของกลุ่มตัวอย่างเท่ากับ 16 อายุเฉลี่ยประชากรเท่ากับ 18 แต่การคาดคะเนก็ไม่มีอคติ เพราะค่าเฉลี่ยที่คาดหวังเพื่อใช้ในการคาดคะเนประชากรเท่ากับค่าเฉลี่ยของประชากร ในทางตรงกันข้าม วิธีการสุ่มตัวอย่างที่มีอคติอาจจะทำให้เกิดกลุ่มตัวอย่างกลุ่มใดกลุ่มหนึ่ง ที่ให้ค่าคาดคะเนที่ถูกต้องได้ ฉะนั้น จึงแสดงให้เห็นว่า คำว่า เกียรติคุณหรืออคติในความหมายเชิงสถิติ มีนัยที่แสดงถึงการให้ผลลัพธ์ที่ถูกต้องเป็นส่วนใหญ่ มิใช่หมายถึงความถูกต้องของกลุ่มตัวอย่างเพียงกลุ่มเดียวเท่านั้น

จากการแจกแจงค่าของกลุ่มตัวอย่างในแบบตัวอย่างง่าย ๆ ดังกล่าวแล้ว หากพิจารณาต่อไปจะพบว่า แบบตัวอย่างใดแสดงให้เห็นหลักการในการสุ่มตัวอย่างอีกประการหนึ่ง กล่าวคือแม้ว่าวิธีการสุ่มตัวอย่างจะเป็นไปตามโฉลกและไม่มีอคติ ค่าที่ใช้คาดคะเนค่าประชากรหาได้เหมือนกันไม่ แต่ละกลุ่มตัวอย่างให้ค่าแตกต่างกันไป และแตกต่างกันไปจากค่าของประชากรที่ต้องการจะคาดคะเนด้วย ในทางปฏิบัติ การคาดคะเนค่าของประชากรใช้ผลลัพธ์ที่ได้จากค่าของกลุ่มตัวอย่างเพียงกลุ่มเดียว เมื่อความจริงตามทฤษฎีปรากฏว่ากลุ่มตัวอย่างแต่ละกลุ่มให้ผลลัพธ์เพื่อนำไปใช้คาดคะเนค่าของประชากรแตกต่างกัน ทำให้รู้สึกว่าการสุ่มตัวอย่างออกจะเป็นวิธีการที่ไม่ค่อยจะน่าเชื่อถือ แต่ในทางเป็นจริงแล้วหาเป็นเช่นนั้นไม่ เพราะเหตุที่เราสามารถทำการคาดคะเนจากค่าของกลุ่มตัวอย่างกลุ่มหนึ่งกลุ่มใด

ที่เลือกมาแบบนี้ได้ว่า มีความแตกต่างเกิดขึ้นในกลุ่มตัวอย่างเหล่านี้มากน้อยเพียงใด ความแตกต่างที่บังเกิดขึ้นระหว่างกลุ่มตัวอย่างเหล่านี้เรียกว่า *ความแปรปรวนของกลุ่มตัวอย่าง* (sampling fluctuation) ฉะนั้น สิ่งที่ต้องการก็คือ มาตรการบางประการที่จะใช้วัดความผันแปรของค่าซึ่งแตกต่างกันจากกลุ่มหนึ่งไปยังอีกกลุ่มหนึ่ง นั่นคือ มาตรการที่จะใช้วัดความกระจายของการแจกแจงค่าของกลุ่มตัวอย่าง โดยปกติแล้วการวัดความกระจายของการแจกแจงค่าใด ๆ นั้น ในทางสถิติใช้ค่า *เบี่ยงเบนมาตรฐาน* (standard deviation) หรือกำลังสองของค่าเบี่ยงเบนมาตรฐานซึ่งเรียกว่า *ความแปรปรวน* (variance) เป็นเครื่องมือวัด

ความเบี่ยงเบนมาตรฐานของการแจกแจงใด ๆ นั้น อาจจะสามารถหาได้โดยสูตรนี้

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

s = ความเบี่ยงเบนมาตรฐาน

x_i = ค่าของเทอมที่ 1 ถึงเทอมที่ n

\bar{x} = ค่าเฉลี่ยของ n เทอม

จากแบบตัวอย่างเรื่องอายุ เราคำนวณความเบี่ยงเบนมาตรฐานของการแจกแจงค่าที่ได้จากกลุ่มตัวอย่าง (sampling distribution) ได้ดังนี้

ค่าเฉลี่ยของกลุ่มตัวอย่างแต่ละกลุ่ม = 16, 16.5, 18.5, 17.5, 19.5, 20

กลุ่มตัวอย่างทั้งหมด 6 กลุ่ม $n = 6$

ความเบี่ยงเบนมาตรฐาน

$$= \sqrt{\frac{(16-18)^2 + (16.5-18)^2 + (18.5-18)^2 + (17.5-18)^2 + (19.5-18)^2 + (20-18)^2}{6}}$$

$$= \sqrt{2.16}$$

$$\approx 1.47$$

ความเบี่ยงเบนมาตรฐานของการแจกแจงค่าที่ได้จากกลุ่มตัวอย่างนี้เรียกกันว่า *ความผิดพลาดของค่าเฉลี่ย* (standard error of the mean) หรือ *ความผิดพลาดของกลุ่มตัว*

อย่าง (sampling error) ซึ่งเป็นมาตรการที่จัดว่าเหมาะสมในการใช้วัดความผันแปรของค่าที่ได้จากกลุ่มตัวอย่างต่าง ๆ กล่าวคือ ใช้วัดหาความถูกต้องที่น่าจะเป็นไปได้ หรือความแม่นยำ (precision) การคาดคะเนของประชากรจากกลุ่มตัวอย่างกลุ่มใดกลุ่มหนึ่งก็ที่ได้กล่าวมาแล้ว

ความผิดพลาดมาตรฐานของค่าเฉลี่ยจากกลุ่มตัวอย่างทั้งหมด อาจจะสามารถหาได้จากสูตรข้างล่างนี้ และสามารถคาดคะเนได้จากกลุ่มตัวอย่างใดกลุ่มหนึ่งด้วย

$$S.E._x = \sqrt{\frac{b^2}{n} \cdot \frac{N-n}{N-1}}$$

b = ความเบี่ยงเบนมาตรฐานของประชากร

N = จำนวนหน่วยในประชากร หรือขนาดของประชากร

n = จำนวนหน่วยในกลุ่มตัวอย่าง หรือขนาดกลุ่มตัวอย่าง

$\sqrt{\frac{N-n}{N-1}}$ เรียกว่า the finite population correction (fpc.) หากว่ากลุ่มตัวอย่างมีขนาดใหญ่เมื่อเปรียบเทียบกับขนาดของประชากร ค่านี้ก็จะใกล้เคียงตัดทิ้งเสียได้ หากประชากรทั้งหมดรวมกันเป็นกลุ่มตัวอย่าง ค่านี้ก็จะเท่ากับศูนย์ จึงไม่มีความเบี่ยงเบนระหว่างกลุ่มตัวอย่างเกิดขึ้น

ถ้าพิจารณาสูตรการหาความผิดพลาดมาตรฐาน $S.E._x = \sqrt{\frac{b^2}{n}}$ หรือ $= \frac{b}{\sqrt{n}}$

โดยไม่พิจารณาค่า $\sqrt{\frac{N-n}{N-1}}$ เราจะเห็นความผิดพลาดมาตรฐานของค่าเฉลี่ยจากกลุ่มตัวอย่างนั้นขึ้นอยู่กับปัจจัย 2 ประการ คือ

1. ขนาดของกลุ่มตัวอย่างหรือค่าของ n จากสามัญสำนึกและประสบการณ์เราพบว่า ความแม่นยำในการคาดคะเนโดยใช้กลุ่มตัวอย่างขนาดใหญ่ มีมากกว่ากลุ่มตัวอย่างขนาดเล็ก และจากสูตรก็จะเห็นว่า ความผิดพลาดมาตรฐานผันแปรเป็นส่วนผกผันกับค่าของกรณฑ์สองของขนาดของกลุ่มตัวอย่าง ซึ่งหมายความว่า ถ้าหากจะลดความผิดพลาดมาตรฐานลงครึ่งหนึ่ง จำเป็นต้องเพิ่มขนาดของกลุ่มตัวอย่างขึ้นเป็น 4 เท่า

2. ลักษณะของประชากรหรือค่าของ 6 คือความเบี่ยงเบนมาตรฐานของกลุ่มประชากร โดยสามัญสำนึก หากค่าแต่ละหน่วยของประชากรแตกต่างกันมาก คือมีลักษณะเป็นวิวิธพันธ์ (heterogeneity) ค่าของกลุ่มตัวอย่างจะใช้ค่าคาดคะเนค่าของประชากรได้แม่นยำ **ตรงน้อยกว่า** ลักษณะของประชากรที่มีความเอกพันธ์ (homogeneity) ฉะนั้น หากแต่ละหน่วยของประชากรมีพิสัยแห่งความแตกต่างกันกว้าง จึงดูออกเป็นการเสี่ยงอยู่มากที่จะใช้ค่าของกลุ่มตัวอย่างคาดคะเนค่าของประชากร หากประชากรมีลักษณะดังกล่าว ลักษณะอันนี้จะแสดงออกมาให้เห็นในการแจกแจงค่าของกลุ่มตัวอย่าง คือ การแจกแจงจะมีการกระจายกว้างขวางด้วย นั่นคือ ค่าของความผิดพลาดมาตรฐานจะสูง ฉะนั้น เมื่อกลุ่มตัวอย่างให้ค่าความผิดพลาดมาตรฐานสูงย่อมแสดงให้เห็นว่า ความเบี่ยงเบนมาตรฐานในกลุ่มประชากรย่อมจะสูงด้วย กล่าวคือ แต่ละหน่วยประชากรมีลักษณะแตกต่างกันมาก

ตามที่เคยกล่าวเน้นไว้แล้วว่า โดยปกติเราไม่ทราบค่าของประชากร ฉะนั้น ค่าความเบี่ยงเบนมาตรฐานของประชากรเราก็ไม่ทราบด้วย เราแก้ไขปัญหานี้ได้โดยวิธีเดียวกัน โดยการใช้ค่าความเบี่ยงเบนมาตรฐานที่คำนวณได้จากกลุ่มตัวอย่าง แทนค่าเบี่ยงเบนมาตรฐานของกลุ่มประชากร แล้วทำการคาดคะเนค่าของความผิดพลาดมาตรฐาน จากสูตร $S.E. = \sqrt{\frac{s^2}{n}}$ ค่าที่ได้นี้เราเรียกว่า ค่าคาดคะเนความผิดพลาดมาตรฐาน (estimated standard error)

สมมติว่า เราทำการสุ่มตัวอย่างนักเรียน 150 คน จากกลุ่มนักเรียนทั้งหมด 3,000 คน เมื่อทำการคำนวณหาค่าต่าง ๆ แล้วพบว่า อายุเฉลี่ยของนักเรียนในกลุ่มตัวอย่าง 150 คน = 21 และความเบี่ยงเบนมาตรฐาน = 12 นั่นคือ $\bar{x} = 21$ $n = 150$ $s = 12$

เพราะฉะนั้น ค่าคาดคะเนความผิดพลาดมาตรฐาน $Est. S.E._{\bar{x}} = \sqrt{\frac{144}{150}} = 0.98$ ความผิดพลาดมาตรฐานนี้ แสดงให้ทราบโดยเฉลี่ยว่า ความเบี่ยงเบนของค่าคาดคะเนที่อาจเป็นไปได้ทุกค่าที่ผิดพลาดไปจากค่าเฉลี่ยของมัน

เมื่อเราทราบค่าต่าง ๆ เหล่านี้จากกลุ่มตัวอย่างที่ได้แล้ว การนำเอาค่าคาดคะเนนี้ไปใช้ประกอบกับทฤษฎีเส้นโค้งปกติ (normal curve) จะทำให้เราคาดคะเนค่าของประชากร

ได้ ตามทฤษฎีเส้นโค้งปกติ ข้อมูลหรือค่าของข้อมูลชุดใดที่มีลักษณะของการแจกแจงเป็นเส้นโค้งปกติ คือการกระจายของข้อมูลชุดนั้น หรือค่าที่กล่าวถึงนั้น มีรูปการกระจายเป็นเส้นโค้งแบบรูประฆังคว่ำ เราจะพบว่า ประมาณ 68% ของมวลประชากร หรือของข้อมูลชุดนั้น จะมีค่าที่กล่าวถึงอยู่ในระหว่าง ค่าที่สูงกว่าและค่าที่ต่ำกว่า ค่าเฉลี่ยของข้อมูลชุดนั้นหนึ่งหน่วยของความเบี่ยงเบนมาตรฐาน และ 95% ของมวลประชากรจะมีค่าที่กล่าวถึงอยู่ในระหว่าง ค่าเฉลี่ย ± 1.96 หน่วยความเบี่ยงเบนมาตรฐาน หรือ 99% จะมีค่าอยู่ในระหว่าง ค่าเฉลี่ย ± 2.58 หน่วยความเบี่ยงเบนมาตรฐาน ฉะนั้น ข้อแถลงเกี่ยวกับประชากรโดยใช้ค่าคาดคะเนจากกลุ่มตัวอย่างจึงกระทำในลักษณะต่อไปนี้ หากว่าค่าเฉลี่ยของกลุ่มตัวอย่างเป็นการคาดคะเนที่ไม่มีเเตียงนั้น สำหรับค่าเฉลี่ยของประชากร (μ) แล้ว จะทำให้มีความมั่นใจในระดับ 68% ว่า ค่าเฉลี่ยของประชากรจะอยู่ในระหว่างค่าเฉลี่ยของกลุ่มตัวอย่างลบด้วยหนึ่งหน่วยของค่าความผิดพลาดมาตรฐาน กับค่าเฉลี่ยของกลุ่มตัวอย่างบวกด้วยหนึ่งหน่วยความผิดพลาดมาตรฐานกล่าวคือ $\bar{x} \pm 6/\sqrt{n}$ ในระดับ 95% ช่วงของการคาดคะเนจะเป็น $\bar{x} \pm 1.96 \cdot 6/\sqrt{n}$ และเป็น $\bar{x} \pm 2.58 \cdot 6/\sqrt{n}$ ในระดับ 99%

จากตัวอย่างที่สมมติในวรรคที่แล้วว่า $\bar{x} = 21$ $n = 150$ $s = 12$ เมื่อคำนวณดู จะพบว่าในระดับความมั่นใจ 68% อายุเฉลี่ยของประชากรจะอยู่ในช่วง $21 \pm 12/\sqrt{150}$ คือระหว่าง 20.02—21.98 เมื่อเพิ่มระดับความมั่นใจให้สูงขึ้นเป็น 95% ช่วงการคาดคะเนย่อมจะต้องกว้างเป็น 19.08—22.92 เมื่อเพิ่มเป็น 99% การคาดคะเนจะอยู่ระหว่าง 18.47—23.53 ในทางปฏิบัติมักจะกล่าวถึงระดับความมั่นใจ 95% ว่าเป็นระดับ 5% และเป็นระดับที่ใช้กันอยู่โดยทั่วไป การเลือกระดับความมั่นใจนั้นอยู่กับการตัดสินใจเลือกของผู้ทำการวิจัยเองว่าในการวิจัยของเขานั้นอะไรเป็นสิ่งสำคัญยิ่งกว่ากัน ระหว่างช่วงแห่งความผิดพลาดแคบ ๆ กับโอกาสที่จะมีความถูกต้องสูง เราไม่สามารถจะเลือกได้ทั้งสองอย่างในเวลาเดียวกัน ทั้งช่วงความผิดพลาดแคบและความน่าจะเป็นไปไม่ได้สูง ถ้าหากเราต้องการจะแถลงด้วยความมั่นใจอย่างยิ่งว่า ค่าของประชากรจะอยู่ในช่วงที่กล่าว เราก็ใช้ช่วงในระดับความมั่นใจ 1%

มากกว่าจะใช้ในระดับ 5% เพราะในระดับนี้มีช่วงการคาดคะเนกว้างกว่า แต่ถ้าหากเรามีความปรารถนาจะให้มีความผิดพลาดรอบ ๆ ค่าที่คาดคะเนน้อย ความมั่นใจที่แสดงเช่นนั้น ก็ย่อมลดน้อยลงตามส่วน

จากข้อความในสองวรรคที่ผ่านมา ผู้อ่านจะพบว่า มีคำใหม่ที่ไม่คุ้นกับสายตา อยู่ 2 คำ คือ ระดับแห่งความเชื่อมั่น หรือระดับของความมั่นใจ และช่วงแห่งความมั่นใจ หรือช่วงของการคาดคะเน จึงขอทำความเข้าใจเกี่ยวกับคำสองคำนี้ไว้ ณ ที่นี้

ระดับแห่งความเชื่อมั่น (Confidence Limits) และช่วงแห่งความมั่นใจ (Confidence Intervals)

ระดับแห่งความเชื่อมั่น หมายถึง อัตราส่วนความแม่นยำของค่าสถิติที่หาได้จากกลุ่มตัวอย่างกลุ่มหนึ่ง ในการสำรวจโดยกลุ่มตัวอย่างนั้น เราสุ่มกลุ่มตัวอย่างขึ้นมาเพียงกลุ่มเดียว ฉะนั้น ค่าสถิติที่ได้จากกลุ่มตัวอย่างควรจะอยู่ในระดับความเชื่อมั่นที่สูงพอควร ถ้ากำหนดระดับความเชื่อมั่น 95% ก็หมายความว่า การสุ่มตัวอย่างจากประชากรเดียวกันโดยใช้วิธีการอย่างเดียวกัน กลุ่มตัวอย่าง 95 กลุ่มใน 100 กลุ่มจะให้ผลลัพธ์อยู่ในช่วงการคาดคะเนที่คำนวณได้ อีก 5 กลุ่มตัวอย่าง จะให้ผลลัพธ์ผิดพลาดไปหรืออาจจะแปลงง่าย ๆ ว่า การคาดคะเนค่าของประชากรโดยวิธีการสุ่มตัวอย่างตามขบวนการที่กล่าวแล้ว ให้ความหวัง 95% ว่าค่าของประชากรที่ต้องการทราบจะอยู่ในช่วงที่คำนวณได้ ฉะนั้นการคาดคะเนคู่ออกจะมีความแม่นยำมากพอๆ หากใช้ระดับความมั่นใจ 99% ก็หมายความว่า มีหวังถูก 99 ครั้ง และผิดเพียง 1 ครั้ง

ส่วนช่วงแห่งความมั่นใจนั้น หมายถึง ช่วงที่เรายอมให้ค่าสถิติที่คำนวณได้มีความคลาดเคลื่อนไปเป็นจำนวนเท่าใด ทั้งนี้เพื่อช่วยให้ค่าที่คำนวณได้นี้มีความแม่นยำในการคาดคะเนมากยิ่งขึ้น ตามที่แสดงให้เห็นแล้วว่า แม้การเลือกกลุ่มตัวอย่างจะเป็นไปโดยไม่มีอคติ แต่ค่าที่คำนวณได้จากกลุ่มตัวอย่างนั้น ไม่ให้ค่าสถิติที่เท่ากับ

ค่าของประชากรเลย ฉะนั้น จะใช้ค่าสถิติของกลุ่มตัวอย่างประเมินค่าของประชากรให้ถูกต้องแน่นอนย่อมเป็นสิ่งที่เป็นไปไม่ได้ มีหน้าซ้ำการประเมินค่าของประชากรนั้นเราประเมินจากผลที่ได้จากกลุ่มตัวอย่างเพียงกลุ่มเดียว ดังนั้นในทางปฏิบัติเราจึงมีความหวังโดยยอมให้มีความคลาดเคลื่อนจำนวนหนึ่ง โดยปกติจำนวนของความคลาดเคลื่อนนั้น การวัดทางสถิติกำหนดจำนวนความคลาดเคลื่อนเป็นหน่วยของความคลาดเคลื่อนมาตรฐาน ทั้งนี้เพราะจากทฤษฎีการสุ่มตัวอย่าง เราทราบว่า 68% ของกลุ่มตัวอย่างจะมีค่าสถิติตกอยู่ในช่วงที่สูงกว่าและต่ำกว่าค่าสถิติอันเดียวกับของประชากร อยู่หนึ่งหน่วยความคลาดเคลื่อนมาตรฐาน และ 95% ของกลุ่มตัวอย่างจะมีค่าสถิติอยู่ในช่วงที่สูงกว่าและต่ำกว่าค่าเดียวกันของประชากรอยู่ 1.96 หน่วยความคลาดเคลื่อนมาตรฐานและ 99% อยู่ในช่วง 2.58 หน่วยความคลาดเคลื่อนมาตรฐาน ในการคำนวณบางครั้งอาจจะใช้ 1 หน่วย 2 หน่วย และ 3 หน่วยความคลาดเคลื่อนมาตรฐานแทนจำนวน 1 หน่วย 1.96 และ 2.58 หน่วยโดยประมาณตามลำดับ

ขอให้สังเกตว่า กลุ่มตัวอย่างกลุ่มหนึ่งนั้น เราใช้ความเบี่ยงเบนมาตรฐาน (σ) เป็นเครื่องวัดความคลาดเคลื่อนของค่าเฉลี่ย (\bar{x}) ของกลุ่มตัวอย่าง แต่ค่าเฉลี่ยของค่าเฉลี่ยจากกลุ่มตัวอย่างที่อาจจะเป็นไปได้ทั้งหมดรวมกัน ($\bar{\bar{x}}$) หรือที่เรียก mean of the sample means นั้น เรามีความคลาดเคลื่อนหรือความผิดพลาดมาตรฐาน ($\sigma_{\bar{x}}$) เป็นเครื่องวัด หากพิจารณาต่อไปจะพบว่า ความคลาดเคลื่อนมาตรฐาน ($\sigma_{\bar{x}}$) นี้ ย่อมมีค่าน้อยกว่าความเบี่ยงเบนมาตรฐาน (σ) เพราะเป็นความคลาดเคลื่อนของค่าเฉลี่ยของกลุ่มตัวอย่างทุกกลุ่มอีกทีหนึ่ง ฉะนั้น เมื่อเราไม่ทราบค่าของความเบี่ยงเบนมาตรฐานของกลุ่มประชากร เราจึงประเมินค่านี้ได้โดยใช้ความเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างแทนโดยประมาณ ซึ่งจะเป็นผลทำให้ช่วงของความมั่นใจกว้างกว่าความเป็นจริง (เมื่อเราทราบค่าเบี่ยงเบนมาตรฐานของประชากร) เล็กน้อย ดังนั้นทำให้มองได้ว่า เรื่องของสถิตินี้เป็นเรื่องของการประมาณ โดยมีความแม่นยำในระดับที่มีความมั่นใจมากพอสมควร

เพื่อทำความเข้าใจในหลักการและเหตุผลดังกล่าวแล้วให้กระจ่างยิ่งขึ้น โดยพยายามหลีกเลี่ยงการคำนวณที่ยุ่งยากเกินไป ขอให้พิจารณาตัวอย่างที่สมมติขึ้นต่อไปนี้

สมมติว่า เรามีนักเรียน 10 คน เป็น A, B, C, D, E, F, G, H, I และ J สอบวิชาหนึ่งได้คะแนนเป็น 0, 1, 2, 3, 4, 5, 6, 7, 8 และ 9 ตามลำดับ นั่นคือเรามีประชากรสมมติ 10 คน คะแนนเฉลี่ยของประชากรเท่ากับคะแนนทั้งหมดรวมกันหารด้วย $10 = 4.5$

หากเราต้องการจะสุ่มตัวอย่างขึ้นสักกลุ่มหนึ่ง ทำการคำนวณหาค่า เพื่อใช้คาดคะเนค่าของประชากร ลองดูซิว่า จะได้ค่าจากกลุ่มตัวอย่างเป็นอะไรบ้าง เมื่อพิจารณาไปจากกลุ่มตัวอย่างที่มีขนาดเล็กที่สุด กลุ่มตัวอย่างที่จะเป็นไปได้ คือ กลุ่มตัวอย่างที่ประกอบด้วยนักเรียน 2 คน ซึ่งเรียกว่ากลุ่มตัวอย่างที่มีขนาด 2 ตามที่เคยกล่าวแล้ว เราอาจจะสุ่มตัวอย่างได้เป็น A กับ B, A กับ C หรือ B กับ C หรืออะไรก็ได้ ถ้าเป็นกลุ่มตัวอย่างขนาด 4 ก็อาจจะเป็น A, B, C และ D หรือ A, B, C และ E หรืออย่างอื่น ถ้าเราเลือกกลุ่มตัวอย่างขนาด 6 ก็คงเป็นเช่นเดียวกัน อาจจะเป็นใครร่วมกลุ่มกับใคร 6 คน หากเราทำแบบนี้แล้วคำนวณหาคะแนนเฉลี่ยของแต่ละกลุ่มตัวอย่างที่อาจจะเป็นไปได้ทั้งหมด แล้วนำค่าที่ได้มาแจกแจงดู จะทำให้เห็นความสัมพันธ์ระหว่างจำนวนหรือเปอร์เซ็นต์ของกลุ่มตัวอย่างกับช่วงการคาดคะเนค่าของประชากร ขอให้พิจารณาคูตารางการแจกแจงค่าที่คำนวณได้จากตัวอย่างที่สมมติข้างบน ซึ่งแสดงไว้ในหน้า 696

จากตารางการแจกแจงคะแนนเฉลี่ยของกลุ่มตัวอย่างขนาดต่าง ๆ ดังกล่าว จะเห็นว่าแม้กลุ่มตัวอย่างแต่ละกลุ่มจะให้คะแนนเฉลี่ยแตกต่างกันไป แต่การเลือกกลุ่มตัวอย่างเป็นการเลือกแบบไม่มีอคติ นั่นคือ เพราะค่าเฉลี่ยของคะแนนเฉลี่ยจากกลุ่มตัวอย่างที่อาจเป็นไปได้ทั้งหมดมีค่าเป็น 4.5 ซึ่งเท่ากับคะแนนเฉลี่ยของประชากร

หากเราทำการสุ่มตัวอย่างนักเรียนเพียง 2 คน ขึ้นมาทำการศึกษาเพื่อวางนัยทั่วไปเกี่ยวกับประชากรกลุ่มนี้ โดยทำการสุ่มตัวอย่างแบบที่เรียกว่า การสุ่มตัวอย่างแบบง่าย (ซึ่งจะได้กล่าวต่อไป) จะพบว่ามีกลุ่มตัวอย่างอยู่เพียง 5 กลุ่มในกลุ่มตัวอย่างที่อาจจะเป็นไปได้ 45 กลุ่ม หรือ 11% ที่จะให้ผลลัพธ์ใกล้เคียงกับค่าของประชากร คือ ระหว่าง

ตารางแสดงการกระจายของคะแนนเฉลี่ยจากกลุ่มตัวอย่างที่อาจจะเป็นไปได้
 ของประชากรนักเรียนสมมติ 10 คน ซึ่งมีคะแนนเฉลี่ย 4.5

คะแนนเฉลี่ยของกลุ่มตัวอย่าง	จำนวนของกลุ่มตัวอย่างขนาด		
	2	4	6
.5	1		
1.0	1		
1.5 — 1.75	2	2	
2.0 — 2.67	5	10	2
2.75 — 3.25	3	25	10
3.33 — 4.00	8	43	52
4.17 — 4.83	5	50	82
5.00 — 5.67	8	43	52
5.75 — 6.25	3	25	10
6.33 — 7.00	5	10	2
7.25 — 7.5	2	2	
8.0	1		
8.5	1		
จำนวนกลุ่มตัวอย่างที่อาจเป็นไปได้ทั้งหมด	45	210	210
ค่าเฉลี่ยของคะแนนเฉลี่ยที่ได้จากกลุ่มตัวอย่างทั้งหมด	4.5	4.5	4.5
เปอร์เซ็นต์ของกลุ่มตัวอย่างที่มีคะแนนเฉลี่ยมากกว่า 4.00 แต่ไม่น้อยกว่า 5.00	11	24	39
เปอร์เซ็นต์ของกลุ่มตัวอย่างที่มีคะแนนเฉลี่ยมากกว่า 2.67 แต่ไม่น้อยกว่า 6.33	60	89	98

4.17—4.83 คะแนน หรืออาจจะพูดได้อีกอย่างหนึ่งว่า มีกลุ่มตัวอย่างเพียง 5 ใน 45 กลุ่มที่มีลักษณะเป็นตัวแทนที่ใกล้เคียงประชากรมากที่สุด หากเราใช้วิธีการสุ่มตัวอย่างแบบนี้จากประชากรเดียวกันนี้ จะพบว่าเรามีความมั่นใจเพียง 11 % เท่านั้นที่จะกล่าวว่า กลุ่มตัวอย่างที่ได้มีลักษณะเป็นตัวแทนที่ใกล้เคียงกับประชากร หากเราขยายลักษณะความใกล้เคียงดังกล่าวออกไปอีกเล็กน้อย กล่าวคือจาก 4.17 — 4.83 คะแนน ออกไปเป็น 2.75 — 6.25 คะแนน ระดับของความมั่นใจจะเพิ่มขึ้นเป็น 60 % เพราะจะมีกลุ่มตัวอย่างที่ให้ค่าเฉลี่ยอยู่ในช่วงนี้ 26 ใน 45 กลุ่มตัวอย่าง ฉะนั้น จะเห็นว่าถ้าจะประเมินค่าของประชากรแบบนี้ ความแม่นยำในการประเมินจะมีน้อย แต่ในขณะเดียวกันระดับความมั่นใจในความถูกต้องจะสูงขึ้น เพราะโอกาสแห่งความถูกต้องในช่วงดังกล่าวมีอยู่ถึง 26 ใน 45

การแจกแจงค่าเฉลี่ยจะเป็นไปในทำนองเดียวกัน สำหรับกลุ่มตัวอย่างที่ประกอบด้วยนักเรียน 4 คน และ 6 คน แต่ในกลุ่มขนาดดังกล่าว จำนวนกลุ่มตัวอย่างที่จะเป็นไปได้ทั้งหมดมี 210 กลุ่มเท่ากัน* กลุ่มตัวอย่างที่ให้ค่าสถิติที่ใกล้เคียงกับประชากรมาก คือระหว่าง 4.17 — 4.83 คะแนน เพิ่มขึ้นเป็น 24 เปอร์เซ็นต์ และ 39 เปอร์เซ็นต์ตามลำดับ หากขยายช่วงแห่งความมั่นใจให้กว้างออกไปเป็น 2.75 — 6.25 คะแนน ระดับแห่งความมั่นใจที่จะเพิ่มขึ้นเป็น 89 และ 98 เปอร์เซ็นต์ตามลำดับ หรืออาจจะกล่าวได้ว่า สำหรับกลุ่มตัวอย่างที่มีนักเรียน 6 คน หากทำการสุ่มตัวอย่างตามวิธีการดังกล่าว มีกลุ่มตัวอย่างเพียง 4 กลุ่มใน 210 กลุ่ม ที่จะให้ค่าสถิติผิดเพี้ยนไปจากค่าระหว่าง 2.75—6.25 คะแนน

$$* \text{คำนวณได้จากสูตร } C_r^n = \frac{n!}{(n-r)! r!}$$

เมื่อ $n =$ จำนวนประชากร (10)

$r =$ จำนวนประชากรในกลุ่มตัวอย่าง (4) หรือ (6)

$$C_4^{10} = \frac{10!}{(10-4)! 4!} = \frac{10!}{6! 4!}$$

$$= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{6 \times 5 \times 4 \times 3 \times 2 \times 1 \cdot 4 \times 3 \times 2 \times 1} = 210$$

$$C_6^{10} = \frac{10!}{(10-6)! 6!}$$

$$= \frac{10!}{4! 6!} = 210$$

นอกจากนี้ ตารางข้างบนยังแสดงให้เห็นหลักเกณฑ์ที่สำคัญอีกอันหนึ่งในทฤษฎีการสุ่มตัวอย่าง กล่าวคือ ความสัมพันธ์ระหว่างขนาดของกลุ่มตัวอย่างกับระดับแห่งความมั่นใจ เมื่อพิจารณาตู่จะพบว่า หากกำหนดช่วงแห่งความมั่นใจไว้ช่วงหนึ่ง ระดับแห่งความมั่นใจจะเพิ่มขึ้นเมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น กล่าวคือ ณ ช่วงแห่งความมั่นใจระหว่าง 4.17—4.88 คะแนน ตามทฤษฎี เรามีความมั่นใจว่า ค่าสถิติของกลุ่มตัวอย่างจะถูกต้องเพียง 11 เปอร์เซ็นต์ หากเราทำการสุ่มตัวอย่างนักเรียน 2 คนมาทำการศึกษา แต่ความมั่นใจจะเพิ่มขึ้นเป็น 24 และ 39 เปอร์เซ็นต์ หากกลุ่มตัวอย่างของเราประกอบด้วยนักเรียน 4 และ 6 คนตามลำดับ ทั้งนี้แสดงว่า กลุ่มตัวอย่างจะให้ค่าสถิติที่มีความแม่นยำหรือมีลักษณะเป็นตัวแทนประชากรมากยิ่งขึ้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ฉะนั้น โดยทั่วไปเมื่อต้องการให้กลุ่มตัวอย่างมีลักษณะเป็นตัวแทนประชากร ก็ควรใช้กลุ่มตัวอย่างขนาดใหญ่ แต่ขอให้พึงระลึกว่าความแม่นยำหรือความเป็นตัวแทนที่เพิ่มขึ้นตามขนาดของกลุ่มตัวอย่างนั้น ย่อมมีข้อจำกัดอยู่หลายประการ โดยเฉพาะอย่างยิ่ง ก็คือค่าใช้จ่ายรายตัวที่เพิ่มขึ้นในการเก็บข้อมูล นอกจากนี้เรายังพบจากประสบการณ์ในการสำรวจต่างๆ อีกว่า กลุ่มตัวอย่างขนาดใหญ่ที่ได้จากขบวนการเลือกและการสำรวจที่ไม่ระมัดระวังนั้น ให้ผลลัพธ์ถูกต้องน้อยกว่ากลุ่มตัวอย่างที่ได้จากวิธีการที่ได้รับการพิจารณาอย่างรอบคอบ การใช้กลุ่มตัวอย่างขนาดใหญ่จึงมิใช่เครื่องประกันว่า จะได้กลุ่มตัวอย่างที่เป็นตัวแทนประชากรเสมอไป นอกจากนี้การคำนวณหาขนาดของกลุ่มตัวอย่างเพื่อจะใช้ในการสำรวจนั้น ยังขึ้นอยู่กับชนิดของกลุ่มตัวอย่างที่ใช้ในการสำรวจด้วย การอธิบายถึงวิธีหาขนาดของกลุ่มตัวอย่างต้องอาศัยความรู้ทางสถิติและการคำนวณที่ยังยากพอสมควร จึงอยู่นอกขอบเขตที่บทความนี้จะกล่าวถึง ท่านผู้สนใจอาจหาอ่านได้จากหนังสือที่กล่าวถึงทฤษฎีการสุ่มตัวอย่างต่างๆ

ชนิดของกลุ่มตัวอย่าง

การวิจัยโดยใช้วิธีการสำรวจ (survey research) นั้น เราจะพบว่ามียู่ 2 แบบด้วยกัน คือการสำมะโน (census survey) ซึ่งหมายถึงการสำรวจทุกหน่วยประชากรที่กำหนดขอบเขตที่แน่นอนไว้แล้ว กับการสำรวจโดยกลุ่มตัวอย่าง (sampling survey) การทำการ

สำรวจโดยการสุ่มตัวอย่างนั้น ผู้ทำการสำรวจจำเป็นต้องเลือกกลุ่มตัวอย่างขึ้นมากลุ่มหนึ่ง จากมวลประชากรที่กำหนดไว้ แล้วจึงทำการสำรวจเฉพาะหน่วยประชากรที่ถูกเลือกขึ้นมา เป็นกลุ่มตัวอย่างเท่านั้น ส่วนหน่วยอื่น ๆ ของประชากรที่เหลือเราไม่คำนึงถึง ทั้งนี้เพราะ จากทฤษฎีการสุ่มตัวอย่างทำให้เรามีความหวังว่า ข้อมูลที่ได้รับจากส่วนที่เลือกมาเป็นกลุ่ม ตัวอย่างจะแสดงให้เห็นถึงลักษณะหรือคุณสมบัติทั่วไปของประชากรที่เราต้องการศึกษา กลุ่มตัวอย่างที่จะเป็นตัวแทนของประชากรหรือแสดงให้เห็น ลักษณะส่วนรวมของประชากร ซึ่งจะเป็นไปตามทฤษฎีดังกล่าว ย่อมขึ้นอยู่กับวิธีการเลือกกลุ่มตัวอย่างนั้น ๆ

วิธีการเลือกกลุ่มตัวอย่างเพื่อใช้ในการวิจัยทางสังคมศาสตร์ ทำให้เกิดกลุ่มตัวอย่างที่ใช้กันอยู่ทั่วไป ซึ่งพอจะแบ่งออกไปได้เป็น 2 ชนิด คือ

1. กลุ่มตัวอย่างที่น่าจะเป็นไปได้
2. กลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้

กลุ่มตัวอย่างทั้งสองชนิดนี้ ยังประกอบขึ้นด้วยกลุ่มตัวอย่างแบบต่าง ๆ อีกหลายแบบซึ่งจะได้กล่าวต่อไป ในขั้นนี้ขอให้พิจารณาความแตกต่างกันระหว่าง กลุ่มตัวอย่างสองชนิดที่กล่าวแล้วเสียก่อน

กลุ่มตัวอย่างที่น่าจะเป็นไปได้ (probability samples) เป็นกลุ่มตัวอย่างที่ได้มาจากวิธีการเลือกที่เรียกว่า "การสุ่ม" (random selection) หน่วยประชากรที่มีการกำหนดขอบเขตไว้แน่นอน มิใช่เป็นการสุ่มที่เป็นไปตามยถากรรม พบใคร เห็นใครหรืออะไรที่มีลักษณะเป็นหน่วยประชากรที่กำหนดขึ้น ก็เลือกจับเอามาเป็นหน่วยหนึ่งของกลุ่มตัวอย่าง เพื่อให้ได้มาซึ่งกลุ่มตัวอย่างที่มีความน่าจะเป็นไปได้ การสุ่มตัวอย่างจะต้องเป็นไปอย่างมีระเบียบและกฎเกณฑ์ กฎเกณฑ์การสุ่มตัวอย่างที่เรียกว่า random sampling นั้น ตามทฤษฎีกล่าวว่า

1. ต้องเป็นการเลือกที่ทำให้แต่ละหน่วยประชากร (element) มีโอกาสหรือโอกาสในการที่จะถูกเลือกเข้ามาเป็นกลุ่มตัวอย่างเท่ากัน หรือเป็นการเลือกที่สามารถจะคำนวณได้ว่า แต่ละหน่วยประชากรมีโอกาสหรือความน่าจะเป็นไปได้ (probability) เท่าไร ที่จะถูกเลือกเป็นกลุ่มตัวอย่าง

2. ต้องเป็นการเลือกที่ทำให้กลุ่มตัวอย่างที่อาจจะเป็นไปได้ทุกหน่วย มีโอกาสที่จะถูกเลือกเท่ากัน

ในทางปฏิบัติ ปรากฏว่าการที่จะปฏิบัติตามกฎเกณฑ์ทั้งสองข้อนี้เป็นของทำได้ยากมาก นอกจากในกรณีที่มีประชากรขนาดเล็ก แต่ในการวิจัยทางสังคมศาสตร์นั้น ส่วนมากผู้ทำการวิจัยมักจะพบกับประชากรจำนวนมาก ฉะนั้น การปฏิบัติตามกฎข้อหนึ่งข้อใด หรือการประยุกต์กฎข้อหนึ่งข้อใดเข้ากับวิธีการสุ่มก็นับว่าเป็นการสุ่มที่ใช้ได้ ซึ่งจะได้กล่าวต่อไปในการเลือกแบบต่าง ๆ ของกลุ่มตัวอย่าง

ส่วนกลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้ (non-probability samples) นั้นเป็นกลุ่มตัวอย่างที่ได้มาจากการเลือกสรรโดยวิธีการอย่างอื่น ซึ่งมีใช้วิธีการสุ่มตามกฎเกณฑ์ที่กล่าวมาแล้ว

เพื่อให้มองเห็นความแตกต่างของกลุ่มตัวอย่างทั้งสองชนิดอย่างแจ่มชัดยิ่งขึ้น เราควรจะได้พิจารณาถึงวิธีการสุ่มตัวอย่างแบบ random sampling เสียก่อน

วิธีการสุ่มตัวอย่าง เมื่อได้กำหนดหน่วยประชากรหรือหน่วยการวิเคราะห์ และทราบขอบเขตของประชากรแล้ว ปัญหาต่อไปก็ขึ้นอยู่กับข้อที่ว่า ทำอย่างไรจึงจะเลือกกลุ่มตัวอย่างให้เป็นไปตามกฎเกณฑ์ของการสุ่มตัวอย่าง ขอให้พิจารณาคู่ตัวอย่างต่อไปนี้ สมมติว่า เราต้องการกลุ่มตัวอย่างขนาด 2 โดยสุ่มจากประชากรสมมติ 5 คน ซึ่งเป็น A B C D และ E วิธีการสุ่มอาจจะทำได้โดยพิจารณาผสมคู่กันทั้ง 5 เป็นคู่ ๆ คือ AB AC AD AE BC BD BE CD CE และ DE รวม 10 คู่ เขียนชื่อแต่ละคู่ลงบนฉลากหรือเปียกลม ๆ ที่ใช้แทนเงินเวลาเล่นไพ่ เอาเปียหรือฉลากใส่ภาชนะเขย่าคลุกเคล้าให้ผสมกันให้ดี แล้วให้คนบีตทายับขึ้นมาหนึ่งอัน ชื่อคู่ใดปรากฏอยู่บนเปีย สองคนนั้นก็ถูกเลือกเป็นกลุ่มตัวอย่างที่เป็นไปตามกฎเกณฑ์การสุ่ม

การเลือกโดยวิธีดังกล่าว จะเห็นว่าเปียแต่ละอันมีโอกาสจะถูกเลือกเท่า ๆ กัน เปียอันหนึ่งแทนกลุ่มตัวอย่างกลุ่มหนึ่ง และแต่ละคนหรือแต่ละอักษรจะปรากฏอยู่บนเปียทั้ง 10 อัน คนละ 4 ครั้งเท่า ๆ กัน ซึ่งหมายถึงแต่ละหน่วยในมวลประชากรมีโอกาสถูกเลือกเท่า ๆ กัน การเลือกจึงเป็นไปตามกฎเกณฑ์ดังกล่าว กลุ่มตัวอย่างที่ได้จึงเรียกว่า

กลุ่มตัวอย่างที่น่าจะเป็นไปได้ ข้อสำคัญพึงสังเกตว่าในการเลือกแบบนี้ การทำฉลากหรือเบี้ยนั้น ฉลากหรือเบี้ยจำเป็นจะต้องเหมือนกันแบบฝาแฝดเลยทีเดียว และฉลากหรือเบี้ย จำต้องทำให้อยู่ในรูปลักษณะที่จะผสมคลุกเคล้าให้เข้ากันให้ดี มิฉะนั้นจะทำให้การเลือกเกิดมั่วคคิ จะทำให้ได้กลุ่มตัวอย่างเทียบคคิตามที่เคยกล่าวมาแล้ว การใช้ลูกบอลในการออกสลากกินแบ่งเป็นตัวอย่างอันดีสำหรับการสุ่มตัวอย่าง เมื่อพิจารณาวิธีการดังกล่าวจะเห็นว่า การจับชิ้นส่วนเพื่อรับรางวัลต่าง ๆ ตามสถานีโทรทัศน์นั้น มีความห่างไกลจากการสุ่มตามความหมายที่กล่าวมานี้มาก

ในทางทฤษฎี การเลือกสุ่มแต่ละหน่วยประชากร เพื่อให้ได้กลุ่มตัวอย่างที่ต้องการตามกฎเกณฑ์ดังกล่าวแล้ว เป็นวิธีการสุ่มตัวอย่างที่เรียกว่า Random Sampling ซึ่งอาจใช้สุ่มตัวอย่างขนาดเท่าใดก็ได้ ในทางปฏิบัติหากประชากรมีจำนวนมากและกลุ่มตัวอย่างที่ต้องการมีขนาดใหญ่ จะพบว่าเราอาจจะต้องใช้เวลาเกือบตลอดชีวิต เพียงแต่เขียนองค์ประกอบของกลุ่มตัวอย่างที่อาจเป็นไปได้ทั้งหมดลงในฉลาก (ขอให้พิจารณาคูสูตร combination ที่ให้ไว้ท้ายหน้า 697) ฉะนั้น จึงได้มีการคิดประดิษฐ์หาวิธีการสุ่มตัวอย่างแบบอื่น เพื่อให้เป็นไปตามหลักเกณฑ์การสุ่มตัวอย่างขึ้น

ความสำคัญของ “การสุ่ม” ในวิธีการเลือกแต่ละหน่วยของประชากรขึ้นเป็นกลุ่มตัวอย่างนั้น เป็นส่วนที่จำเป็นอย่างยิ่งในการป้องกันการเลือกที่มีอคติ แนวความคิดเกี่ยวกับกลุ่มตัวอย่างที่อธิบายแล้วในตอนต้น วางอยู่บนรากฐานของวิธีการสุ่มดังกล่าวแล้ว หากไม่สามารถกำหนดหรือคำนวณหาโอกาส หรือความน่าจะเป็นไปได้ในการเลือกหน่วยจากประชากรเข้ามาเป็นกลุ่มตัวอย่างก็ย่อมจะนำทฤษฎีและแนวความคิดดังกล่าวมาใช้กับกลุ่มตัวอย่างนั้นไม่ได้ การคำนวณหาความผิดพลาดมาตรฐานก็ย่อมทำไม่ได้ ฉะนั้น การใช้ค่าสถิติจากกลุ่มตัวอย่างคาดคะเนค่าสถิติเกี่ยวกับประชากรก็ไม่สามารถจะประเมินความแม่นยำตรงได้เช่นกัน

สามัญชนโดยทั่วไปมักจะเข้าใจว่า “การสุ่ม” หมายถึงการเตาสุ่มไปตามบุญตามกรรม ปะเหมาเคราะห์ดี ถูกล็อตเตอรี่ก็ได้เป็นเศรษฐีเมืองไทย ตามบทเพลงของสลากกินแบ่ง แต่วิธีการออกสลากล็อตเตอรี่เป็นวิธีการเลือกอย่างหนึ่งที่เป็นไปตามหลักเกณฑ์ของการสุ่มในความหมายดังกล่าว แต่แม้กระนั้นก็ยังมียกคนสงสัยกันอยู่และเคยได้ยินคนพูดกันอยู่

บ้างว่าการออกสลากกินแบ่งมีการทำนายได้ มีการถ่วงให้ออกเลขท้ายเบอร์นั้นเบอร์นี้ ผู้เขียนเคยไปเป็นกรรมการออกสลากกินแบ่งครั้งหนึ่ง มองดูวิธีการแล้วไม่เห็นทางที่จะเป็นไปได้ตามเสียงกล่าวหาเลย เว้นไว้เสียแต่ลูกบอลล์กลม ๆ ที่มีหมายเลขกำกับอยู่ เมื่อใช้ไปนาน ๆ อาจจะสึกหรือทำให้มีน้ำหนักไม่เท่ากัน หากลูกบอลล์มีน้ำหนักไม่เท่ากันก็อาจจะทำให้เกิดความเที่ยงคั่นในการออกสลากได้ ถ้าเป็นกลุ่มตัวอย่างก็จะได้กลุ่มตัวอย่างเที่ยงคั่นที่ดังที่กล่าวมาแล้ว แต่เจ้าหน้าที่ก็พยายามระมัดระวังในเรื่องนี้ จึงได้มีการชั่งลูกบอลล์ทุกครั้ง เพื่อตรวจสอบและให้โอกาสแก่หมายเลขทุกหมายเลขเท่า ๆ กัน ที่จะมีย่อออกมาเป็นเลขรางวัลต่าง ๆ โดยเฉพาะอย่างยิ่ง คือรางวัลที่ 1

การตรวจสอบหรือการคัดเลือกใด ๆ โดยมนุษย์เป็นผู้ดำเนินการนั้น เป็นมาตรการที่ไม่น่าพึงใจนัก เมื่อเปรียบเทียบกับความหมายของการสุ่มที่กล่าวถึงแล้ว คนเลือกมักจะมีควมลำเอียงต่อหน่วยหนึ่งหน่วยใดของประชากรไม่มากนักน้อย อาจจะเป็นอย่างนี้โดยสำนึกและโดยคิดไม่ถึง ยิ่งในวงการสำรวจวิจัยทางสังคมแล้ว ความลำเอียงเกิดขึ้นและเห็นได้ง่ายในการเลือก เช่น ให้นักงานสัมภาษณ์ทำการสุ่มเลือกสัมภาษณ์คนบนถนนทั่วไป พนักงานสัมภาษณ์อาจจะเลือกสัมภาษณ์แต่คนที่อยู่ในสังคมระดับเดียวกัน หรืออาจจะโดยไม่ทันคิดเลือกสัมภาษณ์แต่คนที่ประมาณว่าเป็นพวกปานกลาง ซึ่งทำให้ได้กลุ่มตัวอย่างที่มีความแปรผันน้อยกว่าที่เป็นจริงในประชากร หรือไม่ก็เลือกสัมภาษณ์คนที่แต่งตัวเรียบร้อย หน้าตาสวยงาม หรือให้เลือกสัมภาษณ์แม่บ้านทั่วไป พนักงานสัมภาษณ์อาจเลือกเข้าไปสัมภาษณ์แม่บ้านที่มีบ้านสวย ๆ ไม่มีสุนัขเฝ้าบ้าน หรือบ้านที่อยู่ไม่ไกลนัก หรือไม่ลึกนัก ไม่ยอมเข้าไปสัมภาษณ์บ้านที่รกรุงรัง บ้านที่เห็นว่าสกปรก บ้านที่ใหญ่เกินไปเหล่านี้เป็นต้น

เพื่อเป็นการประกันว่า การสุ่มตัวอย่างนั้นเป็นการสุ่มที่เป็นไปตามโศลกจริง ๆ การสุ่มตัวอย่างจึงเป็นวิธีการเลือกที่เป็นอิสระและปราศจากการเกี่ยวข้อใด ๆ กับการตัดสินใจของผู้เลือกนี้เป็นหลักการอันสำคัญของวิธีการสุ่ม นอกจากนี้ การเตรียมการต่าง ๆ ย่อมต้องใช้ความระมัดระวังเป็นพิเศษ ที่จะไม่ให้มีการเลือกที่ร่ำมกที่ซึ่งเกิดขึ้นโดยธรรมชาติ เช่น ฉลากใหญ่บ้างเล็กบ้าง เบียงน้ำหนักไม่เท่ากันหรือการผสมผสานปนเปื้อนเข้ากันดีของ

ฉลาก เป็นต้น วิธีการสุ่มที่ใช้กันอยู่โดยทั่วไปและเป็นไปตามหลักการดังกล่าวมีอยู่ 2 แบบด้วยกันคือ

1. วิธีการจับฉลาก หรือวิธีการออกสลากกินแบ่ง (Lottery Method) คือการเขียนชื่อแต่ละหน่วยประชากรลงในฉลาก ลูกบอลล์หรือเบี้ย ที่มีลักษณะเหมือนกันทุกประการและสามารถจะผสมปนเปกลูกแต่ละลูกกลับกันได้ดี ใส่ลงภาชนะ แล้วทำการเลือกหรือหยิบขึ้นมาทีละอันจนได้จำนวนตามที่ต้องการ หากการเลือกเป็นการสุ่มขึ้นมาทีละหน่วยเมื่อจับขึ้นมาอันหนึ่งแล้วใส่กลับลงไปในที่เดิมผสมกันให้ดี แล้วจับอันต่อไปทำเช่นนั้นจนกว่าจะได้ครบขนาดของกลุ่มตัวอย่างที่ต้องการ หากการสุ่มได้อันที่เคยสุ่มมาแล้ว ก็ไม่นับ ใส่กลับที่เดิมเสีย แล้วสุ่มอีก การสุ่มแบบนี้เป็นการเลือกแบบมีการแทนที่ (selection with replacement) ทั้งนี้เพื่อประกันว่า ทุกหน่วยจะมีโอกาสถูกเลือกเท่ากัน ถ้าจับขึ้นมาอันหนึ่งแล้วไม่ใส่กลับไว้ตามเดิม จับอันต่อไปเลย การสุ่มแบบนี้เรียกว่า เป็นการเลือกแบบไม่มีการแทนที่ (selection without replacement) แบบนี้ความน่าจะเป็นไปได้ของแต่ละหน่วยที่จะถูกเลือกจะไม่เท่ากัน แต่ก็สามารถจะคำนวณได้ว่ามีค่าเป็นเท่าไร

2. วิธีใช้ตารางสุ่มตัวเลข (Random Numbers Method) เป็นวิธีการที่คล้ายกับแบบแรกแต่แทนที่จะทำฉลากให้เสียเวลา นักสถิติได้จัดทำตารางตัวเลขชุดหนึ่งขึ้นประกอบด้วยตัวเลขที่ได้มาจากการสุ่ม เรียงไว้เป็นแถวและเป็นสดมภ์ เพื่อสะดวกในการสุ่มตารางสุ่มตัวเลขมีเลข 0—9 เรียงสุ่มไว้ ลูกกลิ้งปะปนกันเป็นอย่างดี ไม่มีระเบียบในการเรียงและตัวเลขทุกตัวจะปรากฏอยู่ในตารางสุ่มเป็นจำนวนเกือบเท่า ๆ กัน ตารางสุ่มตัวเลขมีหลายแบบ แต่ละแบบมีวิธีใช้แตกต่างกัน หนังสือที่เกี่ยวกับวิชาสถิติจะมีตารางสุ่มตัวเลข (random numbers tables) ไว้แทบทุกเล่ม พร้อมทั้งคำอธิบายวิธีใช้ซึ่งง่ายและสะดวกแก่การสุ่มโดยใช้ตารางตัวเลขนี้ จำเป็นจะต้องมีบัญชีรายชื่อของประชากร ให้เลขอันดับแต่ละหน่วย เช่นมีประชากร 500 คน ก็ให้เลขเรียงลำดับตั้งแต่ 000 ถึง 499 เลือกสุ่มเลขเริ่มต้นจากตาราง แล้วนับเลขเรียงกันไปทีละ 3 หลัก หากเลขชุดใดเป็นหรืออยู่ในระหว่าง 000—499 เลขชุดนั้นก็ถูกเลือกไว้ หากประชากรหน่วยใดมีเลขอันดับตรงกับเลขชุดที่เลือกไว้ นั้น หน่วยนั้นก็ถูกเลือกเป็นหน่วยหนึ่งในกลุ่มตัวอย่าง ถ้าเลขชุดใดอยู่

นอกเหนือ 000--499 ก็ข้ามเลขชุดนั้นไปเสีย ทำเช่นนั้นเรื่อยจนกว่าจะได้กลุ่มตัวอย่างขนาดที่ต้องการ เช่น 50 คน

ในทางปฏิบัติการเลือกกลุ่มตัวอย่างมักจะสุ่มตัวอย่างไว้เกินขนาดที่ต้องการเล็กน้อย เพื่อสำรองไว้เป็นตัวแทนสำหรับหน่วยที่ไม่สามารถจะค้นหาหรือเก็บข้อมูลได้ เพราะเมื่อเลือกสุ่มตัวอย่างได้แล้ว ต้องพยายามเก็บข้อมูลให้ได้จากหน่วยประชากรในกลุ่มตัวอย่างที่เลือกไว้ จะใช้คนอื่นแทนไม่ได้ มิฉะนั้นหลักการสุ่มจะเสียไป จากประสบการณ์จะพบบ่อย ๆ ว่า ผู้ที่เราเลือกได้ล้มหายตายจากไปเสียแล้ว หรือไม่ก็ย้ายที่อยู่ใหม่หรืออาจจะไม่มีโอกาสได้สัมภาษณ์เลย แม้จะไปพบและนัดหมายก็ครั้งแล้วก็ตาม หากจะต้องสุ่มตัวอย่างใหม่ก็จะทำให้เสียเวลามากไป จึงจำเป็นจะต้องเลือกสำรองไว้ล่วงหน้า แต่การไปสัมภาษณ์เพียงคนเดียวแล้วไม่พบบุคคลที่ต้องการ จึงใช้ตัวสำรองแทนเลขนั้นก็ก็เป็นสิ่งที่ไม่ควรกระทำ

กลุ่มตัวอย่างที่ถูกเลือกขึ้นมาจากประชากรโดยวิธีการดังกล่าวข้างบน เรียกว่ากลุ่มตัวอย่างที่น่าจะเป็นไปได้ (probability samples) ส่วนกลุ่มตัวอย่างที่ได้เลือกมาโดยวิธีการอย่างอื่นซึ่งไม่สามารถจะคำนวณหาโอกาส (chance) หรือความน่าจะเป็นไปได้ของแต่ละหน่วยประชากรที่ถูกเลือกเข้ามาเป็นตัวอย่าง ว่าเท่ากันหรือมีความน่าจะเป็นไปได้เท่าไรนั้น เรียกว่า กลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้ (non-probability samples) เช่น สมมติว่าต้องการจะสำรวจดูว่าข้าราชการโดยทั่วไปในพระนครมีความคิดเห็นอย่างไรเกี่ยวกับการเปลี่ยนอัตราและขึ้นเงินเดือนที่รัฐบาลปรับปรุงใหม่ ผู้ทำการวิจัยก็ออกทำการสัมภาษณ์ความคิดเห็นของราชการตามกระทรวงทบวงกรมต่าง ๆ 500 คน ในการสัมภาษณ์เจ้าหน้าที่จะไปสัมภาษณ์ใครที่ไหนเมื่อไร ไม่มีการเลือกไว้ล่วงหน้า ไปพบใคร ใครให้สัมภาษณ์ก็เก็บบันทึกข้อมูลไว้ ใครไม่ให้สัมภาษณ์ ดิถีธุระ ไม่สนใจ ไม่อยู่ ไม่มาทำงานก็ไม่สัมภาษณ์ผู้นั้นเสีย ทำเช่นนั้นจนได้กลุ่มตัวอย่างครบ 500 คน เราจะพบว่าจากกลุ่มตัวอย่างดังกล่าว เราไม่สามารถคำนวณหาโอกาสของแต่ละหน่วยที่ถูกเลือกได้เลย มีหน้าซ้ำประชากรแต่ละหน่วยมีโอกาสไม่เท่ากันในการเลือก เพราะการเลือกไม่ได้เป็นไปตามกรรมวิธีของการสุ่มตัวอย่างดังกล่าวแล้ว

วิธีการสุ่มตามที่ได้กล่าวแล้ว พอทำให้เห็นความแตกต่างระหว่างชนิดของกลุ่มตัวอย่าง ๒ ชนิด และทำให้วินิจฉัยได้ว่า กลุ่มตัวอย่างเป็นกลุ่มตัวอย่างที่น่าจะเป็นไปได้ หรือเป็นกลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้ แต่วิธีการสุ่มตัวอย่าง มิใช่เป็นเครื่องแสดงว่า กลุ่มตัวอย่างชนิดหนึ่งดีกว่าอีกชนิดหนึ่ง กลุ่มตัวอย่างทั้งสองชนิดมีคุณค่าต่อการวิจัยทางสังคมทั้งคู่ ทั้งขึ้นอยู่กับปัจจัยต่าง ๆ ที่เกี่ยวข้องกับการวิจัย เช่น วัตถุประสงค์ ขอบเขตของการวิจัย กำลังคน เวลา เนื้อหา วิธีการ กำลังเงิน และแหล่งข้อมูล หากสิ่งต่าง ๆ เหล่านี้มีอยู่พร้อมแล้ว ควรทำการศึกษารื่องนั้น ๆ ด้วยกลุ่มตัวอย่างที่น่าจะเป็นไปได้ (probability samples) ทั้งนี้เพราะเหตุว่ากลุ่มตัวอย่างแบบนี้ทำให้ผู้วิจัยสามารถคำนวณหาความผิดพลาดที่เกิดขึ้น สามารถพิจารณาด้วยความมั่นใจว่าเป็นกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรที่ทำการศึกษาหรือไม่ ซึ่งจะทำให้ผู้วิจัยวางนัยทั่วไปเกี่ยวกับประชากรได้ สำหรับกลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้ (non-probability samples) นั้น ในทางปฏิบัติใช้กันอยู่มาก โดยเฉพาะอย่างยิ่งในกรณีที่ไม่มีความจำเป็นจะต้องวางนัยทั่วไปเกี่ยวกับประชากร เช่น การศึกษาเฉพาะกรณี หรือการสำรวจที่ไม่สามารถจะกำหนดขอบเขตของประชากรได้ นอกจากนั้นปัจจัยเกี่ยวกับเวลา กำลังเงิน รวมทั้งข้อมูลที่ต้องการก็เป็นแรงผลักดันอันสำคัญที่ทำให้ผู้ทำการวิจัยเลือกกลุ่มตัวอย่างประเภทนี้

ความเข้าใจเกี่ยวกับการสุ่มตัวอย่างเพื่อการวิจัยจะแจ่มชัดยิ่งขึ้น หากได้ทราบถึงแบบของกลุ่มตัวอย่างชนิดต่าง ๆ ที่แตกแยกออกไปจากกลุ่มตัวอย่าง ๒ ประเภทที่กล่าวแล้ว ในการอธิบายแบบของกลุ่มตัวอย่างต่าง ๆ ซึ่งจะได้กล่าวถึงต่อไปนี้ ได้พยายามหลีกเลี่ยงปัญหาการคำนวณที่ยุงยากที่เกี่ยวข้องกับแต่ละแบบของกลุ่มตัวอย่าง ในขั้นนี้ต้องการให้ผู้อ่านได้ติดตามแบบและวิธีการเลือก เพื่อให้ได้มาซึ่งกลุ่มตัวอย่างดังกล่าวเท่านั้น

กลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้ (non-probability samples) กลุ่มตัวอย่างประเภทนี้ในบางตำรา ก็เรียกว่า model sampling การเลือกกลุ่มตัวอย่างประเภทนี้อาศัยข้อสมมติหรือข้อสันนิษฐานอย่างกว้าง ๆ เกี่ยวกับการกระจายของตัวแปรหรือข้อมูลที่ต้องการในกลุ่มประชากร การเลือกเก็บข้อมูลจึงเป็นไปในลักษณะตามยถากรรม ไม่มีกฎเกณฑ์แน่นอนใน

การเลือก ทั้งนี้เพราะคาดว่าลักษณะข้อมูลสำคัญ ๆ ที่ต้องการ กระจายอยู่เสมอใน กลุ่มประชากร หรือกระจายอยู่อย่างไม่เป็นระเบียบ การเลือกตามยถากรรมก็จะทำให้ได้ ข้อมูลที่มีลักษณะเป็นกรณีทั่ว ๆ ไป หรือแบบฉบับทั่วไปของประชากร การวิจัยทาง วิทยาศาสตร์กายภาพ และวิทยาศาสตร์ชีวภาพใช้กลุ่มตัวอย่างประเภทนี้มาก กลุ่มตัวอย่าง ประเภทนี้แบ่งออกเป็นแบบใหญ่ ๆ ได้ 3 แบบคือ

1. กลุ่มตัวอย่างอุบัติเหตุการณ์ (Accidental Sample) เป็นกลุ่มตัวอย่างที่ได้มา จากการเลือกหน่วยประชากรขึ้นมาตามบุญตามกรรม สักแต่ว่าเป็นหน่วยในกลุ่มประชากร ที่เราต้องการ จะเป็นใครก็ได้ สามารถจะให้ข้อมูลที่ต้องการ ก็เก็บบันทึกข้อมูลจากหน่วย นั้น สมมติว่า ถ้าเราคิดว่าผู้ขับขี่รถยนต์จะให้ความคิดเห็นเกี่ยวกับปัญหาความคับคั่งของ การจราจรในกรุงเทพฯ ได้ดี ผู้ทำการวิจัยก็จะออกคำสั่งให้พนักงานสำรวจออกสัมภาษณ์ คนขับรถรับจ้าง รถบรรทุก รถยนต์ส่วนบุคคลตามที่ต่าง ๆ จนได้ตัวอย่างครบขนาดที่ ต้องการ ใครให้สัมภาษณ์ก็สัมภาษณ์ คนไหนไม่ให้สัมภาษณ์ก็ทิ้งไปเสีย หากคนใหม่ต่อไป กลุ่มตัวอย่างที่ได้จากการที่มีผู้อาสาสมัครให้ข้อมูล ก็จัดอยู่ในประเภทนี้ กลุ่มตัวอย่างอุบัติเหตุ การณ์ คำหนึ่งถึงการที่จะได้ข้อมูลมาวิเคราะห์ศึกษาเป็นประการสำคัญ และไม่ก่ออันตราย ระวังในเรื่องความเป็นตัวแทนของกลุ่มตัวอย่างนัก ในตำราบางเล่มเรียกกลุ่มตัวอย่างแบบ นี้ว่า Haphazard Sample หรือ Fortuitous Sample

2. กลุ่มตัวอย่างสัดส่วน (Quota Sample) เป็นกลุ่มตัวอย่างที่นำเอาสัดส่วน ขององค์ประกอบของประชากรมาเป็นเครื่องพิจารณาด้วย โดยทำการศึกษาหรือประมาณ ส่วนประกอบของประชากรเสียก่อน แล้วจึงเลือกกลุ่มตัวอย่างที่จำลองแบบของประชากร มา เช่นสมมติว่าต้องการจะสำรวจความคิดเห็นของนักศึกษาเกี่ยวกับตัวเลขชาธิการที่จะได้รับการ แต่งตั้งใหม่ ในมหาวิทยาลัยมีอยู่ 4 คณะ มีนักศึกษายู่ 40, 25, 20 และ 15% ตาม ลำดับ หากเราต้องการกลุ่มตัวอย่างขนาด 100 คน ผู้สัมภาษณ์ก็เลือกสัมภาษณ์นักศึกษา คณะละ 40, 25, 20 และ 15 ตามลำดับ หากจะแยกเป็นเพศหญิงเพศชาย หรือนักศึกษา ในชั้นต่าง ๆ นั้นย่อมแล้วแต่วัตถุประสงค์ของการวิจัยและจุดประสงค์ในการวิเคราะห์ข้อมูล

แต่อย่างไรก็ตาม การเลือกกลุ่มตัวอย่างสัดส่วนนั้น จะต้องพิจารณาศึกษาลักษณะประชากรเสียก่อน แล้วจึงเลือกกลุ่มตัวอย่างให้มีลักษณะเป็นแบบจำลองขนาดเล็กของกลุ่มประชากรโดยประมาณ กลุ่มตัวอย่างสัดส่วนใช้กันมากในการวิจัยสำรวจประชามติ การวิจัยตลาดหรือการสำรวจในลักษณะที่คล้ายคลึงกัน

3. กลุ่มตัวอย่างต้องประสงค์ (Purposive Sample) เป็นกลุ่มตัวอย่างที่ขึ้นอยู่กับ การพิจารณาตัดสินใจของผู้ทำการวิจัยเองว่า ลักษณะของกลุ่มคนที่เลือกมาเป็นไปตามวัตถุประสงค์ที่ต้องการ ทั้งนี้ยอมอาศัยการพิจารณาอย่างรอบคอบและกลวิธีในการเลือกหน่วยต่าง ๆ ของผู้ทำการวิจัย เช่น เลือกนักเรียนในชั้นหนึ่ง ซึ่งมีสติปัญญาสูงต่ำ ปานกลางมา เพื่อศึกษาเกี่ยวกับวิธีสอนแบบใหม่ เป็นต้น กลวิธีในการเลือกนั้น บางครั้งอาจจะเป็นการเลือกแต่เฉพาะหน่วยที่ผู้วิจัยพิจารณาเห็นว่ามีลักษณะเป็นกลาง ๆ (typical cases) ในเรื่องที่ทำ การวิจัยนั้น ๆ เช่น เลือกสัปดาห์ใด หรือเดือนใดในรอบปีทำการตรวจบัญชี สำรวจอุบัติเหตุบนถนน ตรวจสอบภาวะของตลาดการค้า โดยพิจารณาว่าเป็นสัปดาห์หรือเดือนที่มี ลักษณะเป็นกลาง ๆ ในเรื่องที่จะศึกษา หรือเลือกศึกษาหมู่บ้านที่มีความเจริญปานกลาง เป็นตัวแทนสังคมชนบท หรือสังคมในเมืองแล้วแต่กรณี

การเลือกกลุ่มตัวอย่างดังกล่าวย่อมต้องอาศัยความรู้ ความชำนาญ และ ประสบการณ์ในเรื่องนั้น ๆ ของผู้เชี่ยวชาญ หรือผู้ที่ทำการวิจัย ฉะนั้น กลุ่มตัวอย่างต้อง ประสงค์ จึงมีชื่อเรียกต่าง ๆ กัน เช่น Expert Choice Sample หรือ Judgement Sample

กลุ่มตัวอย่างที่น่าจะเป็นไปได้ (Probability Samples) กลุ่มตัวอย่างประเภทนี้บางทีเรียกว่า "กลุ่มตัวอย่างที่น่าจะเป็นไปได้ตามอุดมคติ" (ideal probability samples) เพราะเป็นกลุ่ม ตัวอย่างที่วางรากฐานอยู่บนวิธีการทางสถิติ ฉะนั้น จึงไม่มีข้อสมมติเกี่ยวกับการกระจายของ ข้อมูลที่ต้องการในกลุ่มประชากร ข้อสมมติที่ว่าหน่วยประชากรกระจายกันอยู่ตามยถากรรม ในกลุ่มประชากร เมื่อการเลือกเป็นไปในแบบตามบุญตามกรรม ย่อมทำให้ได้กลุ่มตัวอย่าง ที่น่าจะเป็นกลุ่มตัวแทนประชากรนั้น แม้จะเป็นข้อสันนิษฐานที่น่าฟัง แต่การเลือกแต่ละ หน่วยประชากรขึ้นมาเป็นกลุ่มตัวอย่างหาได้เป็นไปตามบุญตามกรรมในความหมายทำนอง

เกี่ยวกับการกระจายของหน่วยประชากรไม่ การเลือกหน่วยประชากรย่อมอาศัยการตัดสินใจของบุคคล ซึ่งย่อมจะต้องมีความลำเอียง หรือปัจจัยอย่างอื่นเกิดขึ้นในการเลือก จึงทำให้การเลือกไม่เป็นไปตามยถากรรมในความหมายดังกล่าว ฉะนั้น กลุ่มตัวอย่างประเภทนี้จึงเกิดขึ้นโดยการนำเอาวิธีการเลือกเชิงวิทยาศาสตร์เข้ามาใช้ กล่าวคือ ใช้วิธีการสุ่มที่มีระเบียบและกฎเกณฑ์ โดยจัดให้แต่ละหน่วยประชากรมีโอกาสที่จะถูกเลือกเท่า ๆ กัน หรือมีโอกาสที่จะถูกเลือกที่สามารถจะคำนวณได้ ทั้งนี้เพื่อเป็นเครื่องประกันว่า การเลือกหน่วยประชากรนั้นเป็นการสุ่ม (random selection) หรือเป็นไปตามบุญตามกรรมจริงๆ มิได้เลือกที่รักมักที่ชัง และเป็นไปตามสัดส่วนและโอกาสของแต่ละหน่วยในกลุ่มประชากร กลุ่มตัวอย่างน่าจะเป็นไปได้แบ่งออกเป็นแบบใหญ่ๆ ได้ 4 แบบ คือ

1. กลุ่มตัวอย่างแบบง่าย (Simple Random Sample) เป็นกลุ่มตัวอย่างที่ได้มาจากการเลือกกลุ่มตัวอย่างตามวิธีการสุ่มดังกล่าวมาแล้ว โดยมีการประกันว่า ทุกหน่วยประชากรมีโอกาสที่จะถูกเลือกเป็นกลุ่มตัวอย่างเท่ากัน หรือมีโอกาสที่จะถูกเลือกเป็นส่วนหนึ่งในจำนวนที่สามารถคำนวณได้

2. กลุ่มตัวอย่างมีระบบ (Systematic Sample) เป็นกลุ่มตัวอย่างที่ได้มาจากการสุ่มอีกแบบหนึ่ง กล่าวคือ ก่อนสุ่มตัวอย่างจำต้องทำบัญชีประชากรขึ้นเสียก่อน แล้วจึงทำการแบ่งประชากรออกเป็นช่วง ๆ ช่วงละเท่า ๆ กัน แล้วทำการสุ่มประชากรในช่วงแรกเพียงช่วงเดียว ประชากรในช่วงต่อ ๆ ไปก็จะถูกเลือกไปด้วยโดยอัตโนมัติ เช่น มีประชากรอยู่ 1000 คน ต้องการเลือกกลุ่มตัวอย่างขนาด 100 แบบ systematic sampling หรือ sampling at regular intervals เราแบ่งประชากร 1000 ออกเป็น 100 ช่วง ช่วงละ 10 คน ให้เลขอันดับประชากรจาก 1—1000 ทำฉลากขึ้นมา 10 ใบ มีเลข 1—10 หรือจะใช้วิธีการอย่างอื่น เลือกสุ่มเลขขึ้นมาหนึ่งตัว สมมติว่าเป็นเลข 7 หน่วยของกลุ่มตัวอย่างในช่วงต่อไปจะเป็นหน่วยเลขที่ 17, 27, 37, 47, จนถึง 997 ซึ่งครบขนาดกลุ่มตัวอย่างหนึ่งร้อยที่ต้องการ

3. กลุ่มตัวอย่างชั้นภูมิ (Stratified Random Sample) เป็นกลุ่มตัวอย่างที่ได้จากการสุ่ม โดยทำการแยกประชากรออกเป็นกลุ่มประชากรย่อยเสียก่อน หรือแยกประชากร

ออกเป็นชั้นภูมิ (stratum) แล้วจึงทำการเลือกกลุ่มหน่วยประชากรมาจากแต่ละชั้นภูมิ โดยวิธีการสุ่มตัวอย่างดังกล่าวแล้ว เมื่อเอาหน่วยของประชากรแต่ละชั้นภูมิมารวมกันเข้าจะได้กลุ่มตัวอย่างชั้นภูมิ การเลือกกลุ่มตัวอย่างจากกลุ่มประชากรย่อยนั้น อาจทำได้โดยเลือกสุ่มในจำนวนที่เป็นปฏิภาคกับจำนวนในกลุ่มประชากรย่อย แบบนี้จะเป็นกลุ่มตัวอย่างที่เรียกว่า Proportional Stratified Sample หากเลือกมาไม่เป็นปฏิภาคกับประชากรในแต่ละชั้น จะได้กลุ่มตัวอย่างที่เรียกว่า Unproportional Stratified Sample เหตุผลที่ใช้ในการแยกชั้นภูมิขึ้นอยู่กับคุณลักษณะหรือสมบัติของประชากรที่เกี่ยวข้องกับเรื่องที่ศึกษา การวิจัยโดยทั่วไป ประชากรที่เราทำการวิจัยมักจะถูกแยกออกเป็นชั้นภูมิตามลักษณะของ เพศ อายุ เชื้อชาติ หรือสัญชาติ รายได้ ระดับการศึกษา ศาสนาหรืออาชีพ ฯลฯ

4. กลุ่มตัวอย่างชุกรวม (Cluster Sample) ในบางกรณีบัญชีรายชื่อของประชากร อาจจะจัดหาได้ยาก หรือได้ทำขึ้นได้โดยเสียค่าใช้จ่ายสูง แต่ปรากฏการณ์ในทางสังคมเราพบว่า มนุษย์อยู่รวมกันเป็นกลุ่มก้อน อาจจะเป็นการอยู่อาศัยตามมาตรฐานของกลุ่ม ทำงานร่วมกันในโรงงาน สถานที่ราชการ สมาคม โรงเรียน ร่วมอาชีพเดียวกัน อาคารสงเคราะห์ นอกจากนี้กลุ่มประชากรยังมีนิวาสสถานที่อยู่เป็นกลุ่ม ตามหมู่บ้าน ตำบล อำเภอ และจังหวัด การอยู่กันเป็นกลุ่มก้อนของประชากรนี้ ทำให้เกิดแนวความคิดที่จะเลือกกลุ่มตัวอย่าง เพื่อจะทำการศึกษาประชากรในเรื่องต่าง ๆ โดยเฉพาะอย่างยิ่งในกรณีที่ไม่สามารถจะหาบัญชีรายชื่อของประชากรได้ กลุ่มตัวอย่างชุกรวมเป็นกลุ่มตัวอย่างที่ได้มาจากการแบ่งประชากรออกเป็นกลุ่ม หรือเป็นชุกรวมเสียก่อน โดยใช้เส้นแบ่งเขตตามภูมิศาสตร์ หรือทางการปกครองหรือโดยวิธีการอย่างอื่น แล้วจึงทำการเลือกกลุ่มต่าง ๆ หรือชุกรวมต่าง ๆ ออกมาตามวิธีการสุ่ม โดยพิจารณาว่ากลุ่มหนึ่งหรือชุกรวมหนึ่งเป็นเพียงหน่วยหนึ่ง เมื่อได้กลุ่มหรือชุกรวมแล้วก็ทำการศึกษาทุกหน่วยประชากรทั้งต้องการในกลุ่มหรือชุกรวมนั้น หากจะทำการศึกษาเฉพาะบางหน่วยของประชากรในแต่ละชุกรวมหรือกลุ่ม จำเป็นจะต้องจัดทำบัญชีรายชื่อของแต่ละหน่วยขึ้น แล้วทำการสุ่มตัวอย่างขึ้นมาเป็นกลุ่มตัวอย่างย่อยของแต่ละชุกรวม กลุ่มตัวอย่างย่อยเมื่อรวมกันจะเป็นกลุ่มตัวอย่างของประชากรทั้งหมด วิธีการสุ่มตัวอย่างชุกรวมนี้จะเห็นว่า เราไม่มีความจำเป็นจะต้องจัดทำบัญชีรายชื่อประชากร

ทั้งหมด แต่จัดทำเฉพาะรายชื่อในกลุ่มหรือชุมชนที่ถูกการสุ่มขึ้นมาแล้วเท่านั้น หากหน่วยของประชากรที่ต้องการในแต่ละชุมชนมีไม่มากเกินไปนัก ก็ทำการเก็บข้อมูลจากทุกหน่วยประชากรที่ต้องการ การจัดทำรายชื่อประชากรก็หมดความจำเป็น สมมติว่าจะทำการสำรวจเกี่ยวกับคณะกรรมการอำเภอ เราก็อาจจะทำการสุ่มอำเภอต่าง ๆ ทั่วประเทศมาสัก 50 อำเภอ แต่ละอำเภอก็มีคณะกรรมการอำเภอประมาณ 10 คน เราก็ทำการเก็บข้อมูลจากกรรมการแต่ละคนจากอำเภอที่สุ่มได้ จะเห็นว่าบัญชีประชากรที่ต้องการ คือบัญชีรายชื่ออำเภอต่าง ๆ ทั่วประเทศเท่านั้น มิใช่บัญชีรายชื่อคณะกรรมการอำเภอทั่วประเทศ

เนื่องจากการสุ่มหากสุ่มตัวอย่างชุมชนนี้ มักจะใช้กันมากในชุมชนที่กระจุกกระจายกันอยู่ตามพื้นที่ต่าง ๆ การแบ่งกลุ่มจึงมักจะใช้ขอบเขตทางภูมิศาสตร์ เส้นเขตถนน แม่น้ำ ลำคลอง หรือเขตการปกครองท้องถิ่นที่เป็นเครื่องแบ่งหรือกำหนดขอบเขตของชุมชน ในบางตำราจึงเรียก กลุ่มตัวอย่างชุมชนเป็นกลุ่มตัวอย่างพื้นที่ (areal sample)

จากกลุ่มตัวอย่างที่น่าจะเป็นไปได้ทั้ง 4 แบบ ที่กล่าวมาแล้ว จะเห็นว่า สองพวกแรกเป็นการสุ่มตัวอย่างเพียงชั้นเดียว คือทำการสุ่มหน่วยประชากรเป็นกลุ่มตัวอย่างเลย แต่ในสองแบบหลังนั้นต้องการแบ่งประชากรทั้งหมดออกเป็นชั้นภูมิ (stratum) เป็นกลุ่มหรือชุมชน (cluster) เสียก่อน แล้วจึงทำการสุ่มหน่วยประชากรขึ้นมาเป็นกลุ่มตัวอย่าง การสุ่มหน่วยประชากรในชั้นที่สองนี้อาจจะทำได้ โดยการเลือกสุ่มอย่างมีระบบ (systematic sampling) หรือ การสุ่มแบบง่าย (random sampling) หากการเลือกไม่เป็นไปตามวิธีการสุ่มตัวอย่าง กลุ่มตัวอย่างที่ได้ก็จะเป็นกลุ่มตัวอย่างที่ไม่น่าจะเป็นไปได้ (non-probability sample) ซึ่งอาจจะเป็นแบบหนึ่งแบบใดก็ตามที่กล่าวไว้แล้ว

นอกจากนี้ยังมีกลุ่มตัวอย่างอีกประเภทหนึ่ง ซึ่งเรียกได้ว่าเป็น “กลุ่มตัวอย่างผสม” (Multi-Stage Sample) การเลือกกลุ่มตัวอย่างแบบนี้มีหลายชั้นหลายตอน เป็นการผสมกันระหว่างการเลือกกลุ่มตัวอย่างแบบง่าย แบบชั้นภูมิ และแบบชุมชน กล่าวคือขั้นแรกทำการเลือกชุมชนขึ้นมาก่อน แล้วทำการแบ่งหน่วยประชากรในแต่ละชุมชนออกเป็นหลาย

ชั้นภูมิ ในแต่ละชั้นภูมิทำการสุ่มหน่วยประชากรซึ่งอาจจะเป็นโดยการแทนที่หรือไม่แทนที่ หรือโดยเป็นปฏิภาคหรือไม่เป็นปฏิภาคกับประชากรย่อยในแต่ละชั้นภูมิ ย่อมขึ้นอยู่กับแผนแบบของการวิจัย (research design) ที่ต้องการ เมื่อรวมกลุ่มตัวอย่างย่อยๆ จากแต่ละชั้นภูมิของแต่ละชุมชนเข้าด้วยกันก็จะได้กลุ่มตัวอย่างของประชากรทั้งหมด

สรุป การสุ่มตัวอย่างเพื่อการวิจัยทางสังคมศาสตร์ เป็นส่วนหนึ่งของระเบียบวิธีการวิจัยที่มีรากฐานอยู่บนทฤษฎีการสุ่มตัวอย่าง แม้การสุ่มตัวอย่างจะเป็นงานของผู้เชี่ยวชาญโดยเฉพาะ แต่หลักเกณฑ์ต่าง ๆ ที่กล่าวแล้วย่อมเป็นเรื่องที่นักวิจัยทุกคนควรจะได้ทราบ เพื่อใช้เป็นแนวทางในการพิจารณาทำแผนแบบการวิจัย การดำเนินการเพื่อให้ได้มาซึ่งกลุ่มตัวอย่างที่มีลักษณะเป็นตัวแทนประชากรที่ต้องการจะศึกษา ซึ่งอาจเป็นทั้งกลุ่มตัวอย่างที่น่าจะเป็นไปได้และกลุ่มตัวอย่างที่น่าจะเป็นไปได้นั้น มีความจำเป็นอย่างยิ่งที่ผู้ทำการวิจัยจะต้องศึกษาพิจารณาคุณลักษณะทั่วไปของประชากรให้ละเอียดถี่ถ้วน ความรอบคอบในการศึกษาเรื่องราวของประชากรล่วงหน้า ย่อมทำให้เกิดแนวความคิดว่าจะทำการเลือกกลุ่มตัวอย่างได้อย่างไร จะใช้กลุ่มตัวอย่างแบบไหน จึงจะเหมาะสมกับโครงการวิจัยที่จะทำและ เป็นไปตามวัตถุประสงค์ที่กำหนดไว้ แบบกลุ่มตัวอย่างแต่ละชนิดย่อมมีข้อได้เปรียบและเสียเปรียบแตกต่างกัน หากได้ทำการเลือกกลุ่มตัวอย่างมาด้วยความพินิจพิเคราะห์ กลุ่มตัวอย่างที่น่าจะเป็นไปได้อันหนึ่งให้ผลการวิจัยที่มีคุณค่า ก็มีอยู่เป็นจำนวนมาก มิใช่ว่านักวิจัยจำเป็นจะต้องผูกมัดตัวเองอยู่กับกลุ่มตัวอย่างที่น่าจะเป็นไปได้อันหนึ่ง

กลุ่มตัวอย่างที่เลือกมาโดยมีความน่าจะเป็นไปได้อันหนึ่ง มีความยุ่งยากเกี่ยวกับการคำนวณและการกำหนดเกี่ยวกับความผิดพลาดที่ย่อมให้เกิด การหาขนาดของกลุ่มตัวอย่าง การทดสอบความถูกต้อง ตลอดจนการจัดทำบัญชีรายชื่อประชากร และการเก็บข้อมูลต่างๆ ยิ่งไปกว่านั้น กลุ่มตัวอย่างประเภทนี้ หากพิจารณาวางแผนแบบไม่ถูกต้อง จะทำให้ได้กลุ่มตัวอย่างที่ให้ผลที่อาจจะทำให้เกิดการเข้าใจผิดในผลของการวิจัยอีกด้วย ฉะนั้น จึงกล่าวได้ว่า กลุ่มตัวอย่างเป็นงานเฉพาะของผู้เชี่ยวชาญ ขอให้สังเกตว่าสังคมมนุษย์ประกอบขึ้น

ด้วยกลุ่มประชากรที่สลับซับซ้อนและมีลักษณะแตกต่างกันอยู่อย่างมากมาย ในโลกมีผู้เชี่ยวชาญทางด้านนี้น้อยมาก แต่นั่นมิใช่หมายความว่าเราจะทำการวิจัยไม่ได้ การศึกษาลักษณะของประชากรที่เราจะทำการศึกษาอย่างถ่องแท้ การศึกษาแบบแผนและวิธีการวิจัยที่ผู้ทำไว้แล้วในกลุ่มประชากรที่มีลักษณะเดียวกันหรือคล้ายคลึงกัน หรือการสำรวจความคิดเห็นจากผู้ชำนาญการหรือผู้ที่มีประสบการณ์ในเรื่องนั้น ๆ เป็นวิธีการที่ทำให้เกิดช่องทางและแนวความคิดที่ถูกต้องในการเลือกกลุ่มตัวอย่างเพื่อการวิจัยทางสังคมศาสตร์ อันจะเป็นการนำไปสู่ผลการวิจัยที่ถูกต้องหรือใกล้ความเป็นจริงมากที่สุดเท่าที่จะเป็นไปได้
