บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การอนุมานทางสถิติสำหรับประชากรอันตะจะดำเนินการเพื่อหาข้อสรุปภายใต้แนวคิดที่เชื่อว่าแต่ละหน่วยประชากรมีค่าของตัวแปรที่สนใจเป็นเพียงค่าคงที่ค่าหนึ่งติดอยู่ ระเบียบวิธีทางสถิติสำหรับศึกษาลักษณะประชากรเช่นนี้โดยเริ่มจากการสำมะโน (Census) อันเป็นการเก็บรวบรวมข้อมูลจากทุกหน่วยประชากรแล้วนำข้อมูลที่ได้ไปทำการคำนวณค่าลักษณะประชากร แต่เนื่องจากข้อจำกัดทางด้านเวลาและทรัพยากร ต่อมาในปี ค.ศ. 1850 จึงได้มีการพัฒนาระเบียบวิธีทางสถิติอีกประเภทหนึ่งขึ้น นั่นคือการสำรวจตัวอย่าง (Sample Survey) ซึ่งเป็นการเก็บข้อมูลบางส่วนจากประชากรเพื่อนำไปทำการประมาณค่าลักษณะประชากร

ในยุคแรกๆ การสำรวจตัวอย่างมีเป้าประสงค์หลักเพียงอย่างเดียวคือต้องการอธิบายสภาพความเป็นไปของข้อมูลจากชุดตัวอย่างในเชิงพรรณนา (Descriptive Inference) ในระยะเวลาต่อมาการใช้ประโยชน์จากการสำรวจตัวอย่างยังมีความต้องการในเชิงวิเคราะห์ (Analytic Inference) เพื่อหาข้อสรุปทั้งในเชิงการตั้งข้อสังเกตเกี่ยวกับค่าลักษณะประชากรหรือในเชิงเปรียบเทียบค่าลักษณะประชากรในกลุ่มย่อยหรือระหว่างกลุ่มประชากรหรือลักษณะของความสัมพันธ์ระหว่างปัจจัยในเชิงสาเหตุ การดำเนินการดังกล่าวนี้จึงจำเป็นต้องใช้กระบวนการทางสถิติที่มีความซับซ้อนมากยิ่งขึ้น ในการจัดทำข้อสรุปในเชิงวิเคราะห์กระบวนการทางสถิติที่นำมาใช้ คือ ช่วงความเชื่อมั่น การทดสอบสมมติฐานหรือการสร้างตัวแบบทางสถิติโดยนำระเบียบวิธีทางสถิติสำหรับประชากรอนันต์มาประยุกต์ ซึ่งข้อมูลที่ใช้ต้องมีลักษณะเป็นตัวอย่างสุ่มที่เป็นอิสระกันและมาจากการแจกแจงแบบเดียวกัน แต่เมื่อพิจารณาการสำรวจตัวอย่างในกรณีการสุ่มแบบไม่ใส่คืนซึ่งการสุ่มหน่วยประชากรขึ้นมาเป็นหน่วยตัวอย่างครั้งแรกมีผลต่อการสุ่มครั้งถัดไปเนื่องจากการสุ่มในครั้งแรกจะมีความน่าจะเป็นที่หน่วยประชากรถูกเลือกเป็นหน่วยตัวอย่างเท่ากับ $1/N$ แต่สำหรับการสุ่มครั้งที่สองจะมีความน่าจะเป็นที่หน่วยประชากรถูกเลือกเป็นหน่วยตัวอย่างเท่ากับ $1/N-1$ นั่นคือชุดตัวอย่างขาดคุณสมบัติของตัวอย่างสุ่มที่เป็นอิสระต่อกัน อีกทั้งยังไม่มีรูปแบบความน่าจะเป็นที่กำกับการเกิดขึ้นสำหรับแต่ละค่าตัวแปรของหน่วยประชากร

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จากประเด็นข้างต้นทำให้เกิดคำถามขึ้นว่าการนำข้อมูลที่มาจากการสำรวจตัวอย่างมาทำการประมาณค่าแบบช่วง ซึ่งเป็นกระบวนการที่ใช้สำหรับการประมาณค่าพารามิเตอร์สำหรับประชากรอนันต์นั้นจะเกิดความผิดพลาดขึ้นมากน้อยเพียงใด ดังนั้นในงานวิจัยชิ้นนี้จึงพยายามที่จะตอบคำถามดังกล่าวเพื่อเป็นแนวทางในการจัดทำข้อสรุปในเชิงวิเคราะห์สำหรับข้อมูลที่มาจากการสำรวจตัวอย่างโดยนำระเบียบวิธีทางสถิติสำหรับประชากรอนันต์มาประยุกต์ อีกทั้งยังเป็นการวางรากฐานที่ถูกต้องชัดเจนในการเรียนการสอนสถิติศาสตร์

1.2 วัตถุประสงค์ของการศึกษา

- ทำความเข้าใจถึงกระบวนการพัฒนาทางสถิติที่เกี่ยวข้อง สำหรับเป็นแนวทางในการสร้างช่วงความเชื่อมั่นให้มีความถูกต้องเหมาะสม และมีความน่าเชื่อถือภายใต้ทฤษฎีการสำรวจตัวอย่าง และทฤษฎีการอนุมานทางสถิติ
- เพื่อศึกษาเปรียบเทียบวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยและสัดส่วนประชากรภายใต้แผนการสุ่มตัวอย่างแบบง่ายทั้งแบบใส่คืนและไม่ใส่คืน โดยใช้การประมาณด้วยการแจกแจงแบบปกติของค่าลักษณะประชากรโดยใช้ความแปรปรวนของตัวประมาณ การประมาณด้วยการแจกแจงแบบปกติของค่าลักษณะประชากรโดยใช้ความแปรปรวนของตัวอย่าง และการประมาณด้วยการแจกแจงแบบปกติเมื่อ n มีขนาดใหญ่ ของ M.E. Thompson
- เพื่อศึกษาเปรียบเทียบการประมาณแบบช่วงเมื่อใช้ตัวประมาณแบบ Simple Random Sampling กับการใช้ตัวประมาณแบบ Stratified Random Sampling ภายใต้แผนการสุ่มแบบมีชั้นภูมิอย่างง่าย

1.3 ขอบเขตการวิจัย

ภายใต้แผนการสุ่มตัวอย่างแบบง่าย

- การแจกแจงของประชากรที่ศึกษา ได้แก่ การแจกแจงแบบปกติ การแจกแจงแบบปัวซง และการแจกแจงแบบเอกซ์โพเนนเชียลสำหรับค่าเฉลี่ยประชากร และการแจกแจงแบบแบร์นูลลีสำหรับสัดส่วนประชากร
- ขนาดประชากรที่ศึกษา เท่ากับ 100 500 1,000 2,000 และ 3,000
- ขนาดตัวอย่างเป็นร้อยละ 1, 3, 5 และ 10 ของประชากร

ภายใต้แผนการสุ่มตัวอย่างแบบมีชั้นภูมิอย่างง่าย

- การแจกแจงของประชากรที่ศึกษา ได้แก่ การแจกแจงแบบปกติ การแจกแจงแบบปัวซง และการแจกแจงแบบเอกซ์โพเนนเชียล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

- กำหนดพารามิเตอร์แต่ละชั้นภูมิมีค่าแตกต่างกัน
- ขนาดประชากรที่ศึกษา เท่ากับ 1,500 และ 3,000
- ขนาดของชั้นภูมิมี 2 ลักษณะ คือ มีขนาดเท่ากันทุกชั้นภูมิ นั่นคือ $N_1 = N_2 = N_3$ และแต่ละชั้นภูมิมีขนาดที่แตกต่างกัน นั่นคือ $N_1 = 2N_2 = N_3$
- ขนาดตัวอย่างเป็นร้อยละ 1, 3, 5 และ 10 ของประชากร

1.4 วิธีดำเนินการวิจัย

- จำลองหน่วยประชากรด้วยโปรแกรม Minitab ซึ่งจะทำการศึกษาในกรณีที่ประชากรมีการแจกแจงแบบแบบปกติ การแจกแจงแบบปัวซง การแจกแจงแบบเอกซ์โพเนนเชียลและการแจกแจงแบบแบร์นูลลี ซึ่งมีขนาดประชากร (N) เท่ากับ 10 ถึง 300 จากนั้นทำการสุ่มตัวอย่างเพื่อให้ได้ชุดตัวอย่างทั้งหมดที่เป็นไปได้ โดยใช้โปรแกรม VB (Visual Basic) ซึ่งมีลักษณะการสุ่มตัวอย่างเป็นแบบไม่ใส่คืน แล้วจึงทำการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ยและสัดส่วนประชากร โดยใช้การประมาณช่วงความเชื่อมั่นด้วยการแจกแจงแบบปกติโดยใช้ความแปรปรวนของตัวอย่างและการประมาณช่วงความเชื่อมั่นด้วยการแจกแจงแบบปกติโดยใช้ความแปรปรวนของตัวประมาณ และทำการวัดความคลาดเคลื่อนของการประมาณ โดยพิจารณาจากค่าครอบคลุมความน่าจะเป็นเทียบกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยการศึกษาครั้งนี้กำหนดระดับความเชื่อมั่น เท่ากับ 0.95 เพื่อตรวจสอบความเหมาะสมของช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยและสัดส่วนกรณีที่ขนาดประชากรเท่ากับ 10 ถึง 300

- จำลองหน่วยประชากรให้มีการแจกแจงแบบแบบปกติ การแจกแจงแบบปัวซง การแจกแจงแบบเอกซ์โพเนนเชียลและการแจกแจงแบบแบร์นูลลี ซึ่งมีขนาดประชากร (N) เท่ากับ 100 500 1,000 2,000 และ 3,000 เนื่องจากข้อจำกัดของโปรแกรม VB ที่ไม่สามารถสุ่มชุดตัวอย่างทั้งหมดที่เป็นไปได้ในกรณีที่ประชากรมีขนาดใหญ่ได้ ดังนั้นจึงสุ่มตัวอย่าง 1,000 รอบโดยใช้โปรแกรม Minitab ซึ่งมีลักษณะการสุ่มตัวอย่างทั้งแบบใส่คืนและไม่ใส่คืน กำหนดขนาดตัวอย่างในรูปร้อยละของประชากร 4 ระดับ คือ 1, 3, 5 และ 10 จากนั้นทำการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ยและสัดส่วนประชากร โดยใช้การประมาณช่วงความเชื่อมั่นด้วยการแจกแจงแบบปกติโดยใช้ความแปรปรวนของตัวอย่าง การประมาณช่วงความเชื่อมั่นด้วยการแจกแจงแบบปกติโดยใช้ความแปรปรวนของตัวประมาณและการประมาณช่วงความเชื่อมั่นด้วยการแจกแจงแบบปกติเมื่อ n มีขนาดใหญ่ของ M.E. Thompson จากนั้นทำการวัดความคลาดเคลื่อนของการประมาณ โดยพิจารณาจากค่าครอบคลุมความน่าจะเป็นเทียบกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยการศึกษาครั้งนี้

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

กำหนดระดับความเชื่อมั่น เท่ากับ 0.95 เพื่อตรวจสอบความเหมาะสมของช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยและสัดส่วนกรณีที่ประชากรมีขนาดใหญ่

- จำลองข้อมูลของหน่วยประชากรซึ่งมีการแจกแจงแบบปกติ การแจกแจงแบบปัวซง และการแจกแจงแบบเอกซ์โพเนนเชียล โดยกำหนดให้จำนวนชั้นภูมิมีค่าเท่ากับ 3 มีการกำหนดพารามิเตอร์แต่ละชั้นภูมิมีค่าแตกต่างกัน นั่นคือ $\theta_1 = 2\theta_2 = 4\theta_3$ กำหนดขนาดประชากรเท่ากับ 1,500 และ 3,000 โดยขนาดของชั้นภูมิมี 2 ลักษณะ คือ มีขนาดเท่ากันทุกชั้นภูมิ นั่นคือ $N_1 = N_2 = N_3$ และแต่ละชั้นภูมิมีขนาดที่แตกต่างกัน นั่นคือ $N_1 = 2N_2 = N_3$ เมื่อได้ข้อมูลแล้วจากนั้นทำการสุ่มตัวอย่าง 1,000 รอบโดยใช้โปรแกรม Minitab โดยมีลักษณะการสุ่มตัวอย่างเป็นแบบไม่ใส่คืน กำหนดขนาดตัวอย่างในรูปร้อยละของประชากร 4 ระดับ คือ 1, 3, 5 และ 10 โดยขนาดตัวอย่างชั้นภูมิแบ่งเป็น 2 ลักษณะ คือ สัดส่วนโดยตรงกับขนาดของชั้นภูมิและขนาดเท่ากันทุกชั้นภูมิ แล้วจึงทำการประมาณค่าเฉลี่ยของประชากรโดยใช้การประมาณ 2 วิธีคือ ตัวประมาณแบบ Simple Random Sampling และตัวประมาณแบบ Stratified Random Sampling จากนั้นนำตัวประมาณที่ได้จากทั้ง 2 วิธี ไปประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยประชากร โดยใช้การประมาณด้วยการแจกแจงแบบปกติของค่าลักษณะประชากรโดยใช้ความแปรปรวนของตัวอย่าง และทำการเปรียบเทียบค่าความน่าจะเป็นของการครอบคลุม (Coverage Probability) ที่ได้จากการประมาณทั้งสองวิธีว่าวิธีใดให้ค่าครอบคลุมความน่าจะเป็นใกล้เคียงกับค่าครอบคลุมความน่าจะเป็นที่กำหนดมากกว่ากัน โดยการศึกษาครั้งนี้กำหนดระดับความเชื่อมั่น เท่ากับ 0.95 เพื่อตรวจสอบความเหมาะสมของช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยภายใต้แผนการสุ่มแบบมีชั้นภูมิอย่างง่าย

1.5 คำจำกัดความ

ประชากร (Population) หมายถึง กลุ่มของคน สัตว์ สิ่งของ วัตถุ กลไก หรืออะไรก็ตามแต่ที่อยู่ภายใต้กรอบความสนใจที่ต้องการศึกษา

ตัวแปรสุ่ม (Random Variable) คือฟังก์ชันที่ใช้อธิบายผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองสุ่มในปริภูมิตัวอย่างให้อยู่ในรูปของจำนวนจริง (Real Number) ให้กับทุกจุดตัวอย่างที่อยู่ภายใต้ปริภูมิตัวอย่างของการทดลองสุ่มนั้นๆ

ฟังก์ชันความน่าจะเป็น (Probability Function) เป็นฟังก์ชันทางคณิตศาสตร์ที่มีคุณสมบัติเฉพาะในอันที่จะนำมาใช้สำหรับอธิบายโอกาสที่จะเกิดขึ้นของเหตุการณ์ต่างๆ ทั้งหมดที่เป็นสมาชิกของปริภูมิตัวอย่างภายใต้ขอบเขตการศึกษา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

พารามิเตอร์ (Parameter) คือ ค่าซึ่งกำกับพฤติกรรมของรูปแบบความน่าจะเป็นซึ่งในกรณีประชากรอนันต์ก็คือ รูปแบบที่กำกับการเกิดขึ้นของหน่วยประชากร ซึ่งหากทราบค่าเหล่านี้ก็จะทำให้เห็นถึงพฤติกรรมโดยรวมของหน่วยประชากรทั้งหมดผ่านรูปแบบความน่าจะเป็นนั้นๆ

ค่าลักษณะประชากร (Population Characteristics) เป็นค่าคุณลักษณะจำเพาะของประชากรที่กำลังศึกษา ซึ่งคำนวณโดยใช้ข้อมูลจากทุกหน่วยประชากรเพื่อแสดงให้ทราบถึงสภาพความเป็นไปของประชากรนั้นๆ

ค่าเฉลี่ยประชากร (Population Mean) เป็นค่าลักษณะประชากรที่ทำให้ทราบว่าลักษณะประชากรที่สนใจโดยเฉลี่ยของแต่ละหน่วยประชากรเป็นเท่าใด

สัดส่วนประชากร (Population) หมายถึง จำนวนหน่วยในประชากรที่มีลักษณะที่สนใจเทียบกับจำนวนหน่วยทั้งหมดในประชากร

ความน่าจะเป็นของการเกิดผลสำเร็จ (Probability of Success) หมายถึง โอกาสที่จะเกิดเหตุการณ์ที่สนใจในการทดลองแต่ละครั้งภายใต้การทดลองสุ่มแบบแบร์นูลลี

ตัวอย่างสุ่ม (Random Sample) หมายถึง ชุดของตัวแปรสุ่ม n ตัว ที่มีการแจกแจงเดียวกันและเป็นอิสระกันที่ได้มาจากประชากรอนันต์หรือประชากรอันตะ

1.6 ประโยชน์ที่จะได้รับจากการศึกษา

- เกิดความเข้าใจที่ชัดเจนในทฤษฎีการอนุมานทางสถิติ และทฤษฎีการสำรวจตัวอย่างเกี่ยวกับการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยและสัดส่วนประชากรในประชากรอันตะจนสามารนำหลักการทางทฤษฎีไปใช้ในทางปฏิบัติได้อย่างถูกต้อง
- เพื่อเป็นแนวทางในการศึกษาระเบียบวิธีทางสถิติสำหรับการจัดทำข้อสรุปในเชิงวิเคราะห์สำหรับประชากรอันตะต่อไป
- เพื่อเป็นรากฐานสำหรับการเรียนการสอนสถิติศาสตร์ทั้งในส่วนของทฤษฎีความน่าจะเป็น ทฤษฎีการอนุมานทางสถิติ และทฤษฎีการสำรวจตัวอย่าง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved