

บทที่ 6
สถิติอนุมาน
(INFERENCE STATISTICS)

วัตถุประสงค์

เมื่อท่านศึกษาเนื้อหาบทที่ 6 โดยละเอียดแล้ว ควรมีความสามารถ ดังนี้

1. บอกความหมายของสถิติอนุมานได้
2. บอกวิธีการที่ใช้ในสถิติอนุมานได้
3. บอกความแตกต่างระหว่าง ประชากร และ กลุ่มตัวอย่างได้
4. บอกความแตกต่างระหว่างค่าพารามิเตอร์ และค่าสถิติได้
5. อธิบายความหมายของทฤษฎีลิมิตกลาง ได้
6. เข้าใจหลักการของวิธีการสุ่มตัวอย่างชนิดต่าง ๆ

เนื้อหา

- 6.1 ความหมายของสถิติอนุมาน
- 6.2 วิธีการที่ใช้ในสถิติอนุมาน
- 6.3 ประชากร และกลุ่มตัวอย่าง
- 6.4 พารามิเตอร์ และ สถิติ
- 6.5 ทฤษฎีลิมิตกลาง (Central Limit Theorem)
- 6.6 วิธีการสุ่มตัวอย่าง

เนื้อหาที่ 6.1 ความหมายของสถิติอนุมาน

สถิติอนุมาน หมายถึง สถิติที่เกี่ยวข้องกับเรื่องดังต่อไปนี้

- 1) การหาความรู้เกี่ยวกับสิ่งของ บุคคล หรือเหตุการณ์จำนวนมาก หรือกลุ่มใหญ่ โดยอาศัยข้อมูลที่ศึกษาได้จากกลุ่มเล็ก ๆ ของสิ่งของ บุคคล หรือเหตุการณ์นั้น ๆ
- 2) การหาข้อสรุปเกี่ยวกับค่าพารามิเตอร์ของประชากร (population parameters) โดยใช้ข้อมูลทางสถิติที่คำนวณจากกลุ่มตัวอย่างที่สุ่มมา (random sample)

เนื้อหาที่ 6.2 วิธีการที่ใช้ในการสถิติอนุมาน

วิธีการที่ใช้ในการอนุมานหรือขยายความเพื่ออ้างอิงไปยังกลุ่มประชากรโดยใช้ข้อมูลที่ได้จากกลุ่มตัวอย่างที่สุ่มมาจากประชากร มี 2 วิธี คือ

1. การประมาณค่า (estimation) เป็นการประมาณค่าหรือหาค่าโดยประมาณของพารามิเตอร์ การประมาณค่าของพารามิเตอร์อาจทำได้ 2 แบบ คือ

ก. การประมาณค่าพารามิเตอร์แบบเป็นจุด เช่นประมาณว่าความสามารถเฉลี่ยในวิชาคณิตศาสตร์ของประชากรนี้คือ 80

ข. การประมาณค่าพารามิเตอร์แบบเป็นช่วง เช่นประมาณว่า ความสามารถเฉลี่ยในวิชาคณิตศาสตร์ของประชากรนี้อยู่ระหว่างช่วงคะแนน 75-90

2. การทดสอบสมมุติฐาน (Hypothesis testing) เป็นการคาดคะเนค่าของประชากรหรือค่าพารามิเตอร์โดยใช้วิธีการทางสถิติทำการทดสอบว่าค่าของประชากรที่คาดคะเนหรือตั้งสมมุติฐานขึ้นมาั้นถูกต้องตามหลักสถิติหรือไม่เช่น ถ้ามีการตั้งสมมุติฐานว่า ความสามารถเฉลี่ยทางคณิตศาสตร์ของประชากรเท่ากับ 80 เราจะใช้วิธีการทางสถิติทำการทดสอบว่าความสามารถเฉลี่ยทางคณิตศาสตร์ของประชากรกลุ่มนี้เท่ากับ 80 หรือไม่

เนื้อหาที่ 6.3 ประชากร และกลุ่มตัวอย่าง

6.3.1 ประชากร

ประชากร หมายถึง กลุ่มหรือข้อมูลทั้งหมดที่ผู้วิจัยจะศึกษาหรืออาจกล่าวได้ว่า ประชากร หมายถึง เซ็ทสากล (Universal set) กลุ่มประชากรอาจมีจำนวนจำกัด (finite) หรือไม่จำกัด (infinite) ก็ได้ เช่น ถ้าเราต้องการศึกษาเกี่ยวกับรายได้เฉลี่ยของคนไทย กลุ่มประชากรของสิ่งที่จะศึกษา คือ “คนไทยทั้งหมด” หรือถ้าต้องการทราบความสามารถเฉลี่ยของนักศึกษาที่เรียนวิชา MR 371 กลุ่มประชากร คือ นักศึกษาที่เรียนวิชา MR 371 ทั้งหมด

6.3.2 กลุ่มตัวอย่าง

กลุ่มตัวอย่าง คือ อนุเซ็ท (subset) ของประชากร กลุ่มตัวอย่างที่ดีต้องเป็นกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรทั้งหมด วิธีการที่จะให้ได้มาซึ่งกลุ่มตัวอย่างที่ดีนั้นอาจทำได้โดยการเลือกตัวอย่างแบบสุ่ม คือเลือกมาอย่างไม่จำเพาะเจาะจงว่าสมาชิกใดในกลุ่มประชากรจะต้องเป็นตัวอย่าง แต่จะให้สมาชิกทุกคนในกลุ่มประชากรมีโอกาสหรือมีค่าความน่าจะเป็นเท่า ๆ กัน ในการได้รับเลือกเป็นกลุ่มตัวอย่าง การเลือกกลุ่มตัวอย่างด้วยวิธีนี้เรียกว่า การเลือกกลุ่มตัวอย่างแบบสุ่ม (random sampling) ดังนั้นกลุ่มตัวอย่างที่ได้มา

สถิติอนุมาน

โดยวิธีสุ่มจึงมีชื่อเรียกว่า กลุ่มตัวอย่างสุ่ม (random sample)

เนื้อหาที่ 6.4 พารามิเตอร์ และสถิติ

6.4.1 พารามิเตอร์

พารามิเตอร์ คือค่าที่แสดงคุณลักษณะของประชากร เช่น ค่าเฉลี่ยและความแปรปรวนของประชากรเราใช้สัญลักษณ์ μ แทนค่าเฉลี่ยของประชากร ส่วนสัญลักษณ์ที่ใช้แทนความแปรปรวนของประชากรคือ σ^2

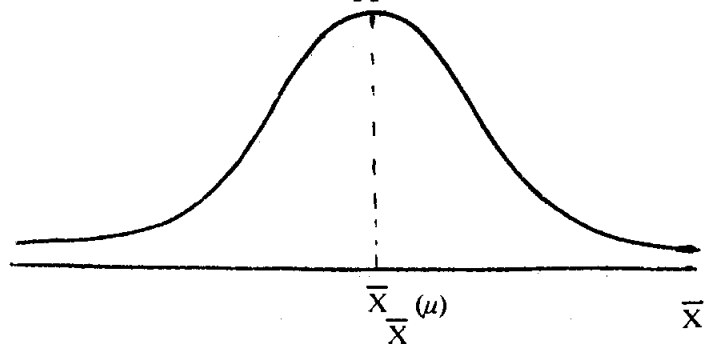
ในการหาค่าพารามิเตอร์ของประชากรนั้น ถ้าประชากรมีจำนวนจำกัด เราสามารถหาได้โดยตรงจากประชากร แต่ถ้าประชากรมีจำนวนไม่จำกัด ค่าพารามิเตอร์ก็จะเป็นค่าทางทฤษฎี ซึ่งไม่สามารถจะหาได้โดยตรง วิธีที่จะหาค่าพารามิเตอร์ได้คือ หาได้โดยการประมาณจากค่าที่ได้จากกลุ่มตัวอย่าง

6.4.2 สถิติ

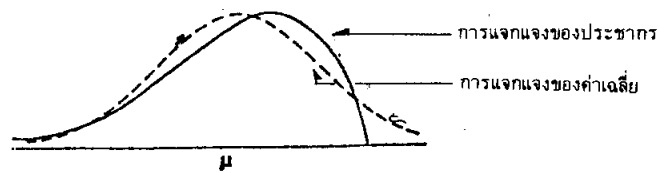
สถิติ คือค่าต่าง ๆ ที่คำนวณได้จากกลุ่มตัวอย่าง เช่น ค่าเฉลี่ย หรือความแปรปรวนของกลุ่มตัวอย่าง เราใช้สัญลักษณ์ \bar{X} แทนค่าเฉลี่ยของกลุ่มตัวอย่าง และ S^2 แทน ความแปรปรวนของกลุ่มตัวอย่าง

เนื้อหาที่ 6.5 ทฤษฎีลิมิตกลาง

ทฤษฎีลิมิตกลาง กล่าวว่า ถ้ากลุ่มตัวอย่างหลาย ๆ กลุ่มถูกสุ่มมาจากประชากรที่มีค่าเฉลี่ยคือ μ และส่วนเบี่ยงมาตรฐาน คือ σ และถ้าจำนวนกลุ่มตัวอย่างที่สุ่มมาของแต่ละกลุ่มมีจำนวนมากกว่า 30 แล้ว ค่าเฉลี่ย (\bar{X}) ของกลุ่มตัวอย่างจะมีการแจกแจงแบบปกติ (normal distribution) ไม่ว่าการแจกแจงของประชากรเดิมจะเป็นแบบใดก็ตาม รูปข้างล่างแสดงให้เห็นการกระจายของค่าเฉลี่ยของกลุ่มตัวอย่างหลาย ๆ กลุ่มที่สุ่มมาจากประชากร ซึ่งอาจจะมีการกระจายเป็นแบบใดก็ได้ (ดูรูป 6.2 และ 6.2 ประกอบ)



รูป 6.1 แสดงการแจกแจงของค่าเฉลี่ยของกลุ่มตัวอย่างที่มีกลุ่มตัวอย่างมากกว่า 30



รูป 6.2 แสดงการแจกแจงของประชากร ซึ่งมีการแจกแจงแบบเบ้ไปทางลบ และการแจกแจงของค่าเฉลี่ย ตามทฤษฎีลิมิตกลาง

เนื้อหาที่ 6.6 วิธีการสุ่มตัวอย่าง

การสุ่มตัวอย่างเป็นวิธีการเลือกตัวอย่างจากประชากรโดยปราศจากอคติหรือการลำเอียง เพื่อให้ได้ตัวอย่างที่เป็นตัวแทนของประชากร และสามารถใช้อำนาจจากกลุ่มตัวอย่างนั้นอ้างอิงหรือสรุปไปสู่ประชากรที่กำหนดไว้ได้อย่างเหมาะสม วิธีการสุ่มตัวอย่างให้เป็นตัวแทนของประชากรที่นิยมใช้กันมี 4 วิธีคือ

- 6.6.1 การสุ่มแบบง่าย (simple random sampling)
- 6.6.2 การสุ่มแบบมีระบบ (systematic sampling)
- 6.6.3 การสุ่มแบบแบ่งชั้น (stratified random sampling)
- 6.6.4 การสุ่มแบบแบ่งกลุ่ม (cluster sampling)

6.6.1 การสุ่มแบบง่าย

การสุ่มตัวอย่างแบบง่าย เป็นวิธีสุ่มตัวอย่างที่หน่วยประชากรทุกหน่วยที่กำหนดไว้ในขอบเขตของการวิจัยมีโอกาสที่จะได้รับการเลือกมาเป็นกลุ่มตัวอย่างเท่า ๆ กัน การสุ่มตัวอย่างแบบง่ายนี้ อาจทำได้โดยเขียนหมายเลขของประชากรทุกหน่วย แล้วทำการสุ่มเลือกกลุ่มตัวอย่างโดยใช้ตารางเลขสุ่ม หรือใช้วิธีจับสลาก เพื่อให้ได้จำนวนกลุ่มตัวอย่างที่ต้องการ อนึ่ง การสุ่มแบบง่ายโดยวิธีจับสลากนั้น เมื่อจับได้หมายเลขใดหมายเลขหนึ่งแล้ว ต้องคืนสลากนั้นกลับไปยังกลุ่มประชากรก่อนแล้วจึงจับสลากในครั้งถัดไป ทั้งนี้เพื่อให้ประชากรทุกคนมีโอกาสที่จะถูกเลือกเท่าเทียมกันในการจับสลากแต่ละครั้ง

6.6.2 การสุ่มแบบมีระบบ

การสุ่มตัวอย่างแบบมีระบบ นิยมใช้ในกรณีที่ประชากรถูกจัดเรียงลำดับไว้ในบัญชีหรือในบัตรข้อมูลเช่น บัญชีรายชื่อนักเรียน บัญชีรายชื่อโรงเรียน เรียงตามเขตการศึกษา บัญชีรายชื่อผู้ใช้โทรศัพท์ ในการสุ่มกลุ่มตัวอย่างแบบมีระบบจะ **ใช้วิธีเลือกข้อมูลรายการ**

แรกจากประชากรซึ่งถูกเรียงลำดับไว้มาโดยวิธีสุ่มแบบง่ายมาเพียงรายการเดียวก่อน จากนั้นจึงเลือกข้อมูลรายการต่อไปโดยนับเป็นช่วง ๆ ละเท่า ๆ กัน เช่นถ้าต้องการสุ่มกลุ่มตัวอย่าง

นักเรียน 100 คน จากนักเรียนกลุ่มประชากร 500 คน ที่มีชื่อเรียงลำดับอยู่ในบัญชี โดยใช้การสุ่มแบบมีระบบก็ทำได้โดยทำการสุ่มนักเรียนคนแรกโดยใช้วิธีการสุ่มแบบง่ายเสียก่อน (สมมุติว่าได้หมายเลข 6) จากนั้นจึงหาจำนวนนักเรียนแต่ละช่วง ซึ่งเท่ากับ 5 ($500/100$) ดังนั้นนักเรียนคนที่ 2 ที่จะได้รับเลือกมาเป็นกลุ่มตัวอย่างคือ หมายเลข 11 ($6+5$) และคนที่ 3 คือหมายเลข 16 ($11+5$) ทำดังนี้เรื่อยไปจนครบจำนวนกลุ่มตัวอย่างที่ต้องการ (100 คน)

6.6.3 การสุ่มแบบแบ่งชั้น

การสุ่มตัวอย่างแบบแบ่งชั้นมีหลักการว่าผู้ที่ทำหน้าที่สุ่มตัวอย่างจะต้องแบ่งประชากรออกเป็นชั้น ๆ เสียก่อน โดยยึดหลักว่า ภายในแต่ละชั้นจะต้องมีลักษณะคล้ายคลึงกันมากที่สุด และ ให้ระหว่างชั้นมีความแตกต่างกันมากที่สุด หรืออาจกล่าวได้ว่า การสุ่มตัวอย่างแบบแบ่งชั้น เป็นการสุ่มตัวอย่างโดยการแบ่งประชากรออกเป็นชั้น ๆ ซึ่งสมาชิกภายในแต่ละชั้นจะมีลักษณะเป็นเอกพันธ์ (homogeneous) และสมาชิกในระหว่างมีลักษณะเป็นวิวิธพันธ์ (heterogeneous) โดยที่สมาชิกหนึ่งเมื่ออยู่ในชั้นใดแล้ว จะไปปรากฏอยู่ในชั้นอื่น ๆ อีกไม่ได้ในการแบ่งชั้นนี้ จะต้องแบ่งให้สอดคล้องกับตัวแปรอิสระที่ต้องการศึกษา เช่น แบ่งตามระดับการศึกษาตามอายุ ตามอาชีพ หลังจากนั้น จึงทำการเลือกสมาชิกภายในแต่ละชั้น ด้วยวิธีการสุ่มแบบง่าย หรือจะใช้วิธีการสุ่มแบบมีระบบภายในแต่ละชั้นก็ได้

6.6.4 การสุ่มแบบแบ่งกลุ่ม

การสุ่มตัวอย่างแบบแบ่งกลุ่ม มีชื่อเรียกอีกชื่อหนึ่งว่าการสุ่มแบบพื้นที่ (area sampling) เป็นวิธีการสุ่มตัวอย่าง โดยการแบ่งประชากรออกเป็นกลุ่มย่อย ๆ เสียก่อน โดยยึดหลักการว่า ในการแบ่งกลุ่มประชากรเป็นกลุ่มย่อย ๆ นี้จะต้องให้ลักษณะภายในแต่ละกลุ่มประชากรย่อยมีความแตกต่างกันมากที่สุดเท่าที่จะเป็นไปได้ ในขณะที่เดียวกันจะต้องให้มีความแตกต่างระหว่างกลุ่มน้อยที่สุดเท่าที่จะเป็นไปได้ (ซึ่งตรงกันข้ามกับหลักการสุ่มตัวอย่างแบบแบ่งชั้น) แล้วจึงใช้วิธีการสุ่มตัวอย่างแบบง่าย สุ่มตัวอย่างแต่ละกลุ่มมาอีกที ในการแบ่งกลุ่มแบบนี้ผู้ต้องการสุ่มตัวอย่างอาจใช้สภาพทางภูมิศาสตร์เช่น ถนน หมู่บ้าน หรือสถานที่ เป็นหลักในการแบ่งกลุ่มก็ได้ การสุ่มตัวอย่างแบบกลุ่มนี้จะประหยัดค่าใช้จ่าย และแรงงานมากกว่าการสุ่มตัวอย่างแบบแบ่งชั้น

สรุปเนื้อหาบทที่ 6

1. สถิติอนุमानเป็นสถิติที่เกี่ยวกับการคาดคะเนหรืออนุमानค่าพารามิเตอร์ของประชากร โดยใช้ข้อมูลทางสถิติที่คำนวณจากกลุ่มตัวอย่างที่สุ่มมา
2. วิธีการที่นิยมใช้ในการคาดคะเนหรืออนุमानค่าของประชากรมี 2 วิธีคือ วิธีการประมาณค่า และวิธีการทดสอบสมมุติฐาน
3. ประชากรหมายถึง กลุ่ม หรือ ข้อมูลทั้งหมดที่ผู้วิจัยต้องการจะศึกษา
4. กลุ่มตัวอย่างคืออนุเซ็ทของประชากร กลุ่มตัวอย่างที่ดีต้องเป็นกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรทั้งหมด
5. พารามิเตอร์ คือค่าที่แสดงคุณลักษณะของประชากร
6. สถิติ คือค่าที่คำนวณได้จากกลุ่มตัวอย่าง
7. ทฤษฎีลิมิตกลาง กล่าวว่่า ถ้ากลุ่มตัวอย่างหลาย ๆ กลุ่มถูกสุ่มมาจากประชากรที่มีค่าเฉลี่ย μ และมีค่าส่วนเบี่ยงเบนมาตรฐาน σ และถ้าจำนวนกลุ่มตัวอย่างที่สุ่มมาของแต่ละกลุ่มมีจำนวนมากกว่า 30 แล้ว ค่าเฉลี่ยของกลุ่มตัวอย่างจะมีการแจกแจงแบบปกติ ไม่ว่าการแจกแจงของประชากรเดิมจะเป็นแบบใดก็ตาม
8. การสุ่มตัวอย่างเป็นวิธีการเลือกตัวอย่างจากประชากรโดยปราศจากอคติ หรือการลำเอียง เพื่อให้ได้ตัวอย่างที่เป็นตัวแทนของประชากร
9. วิธีการสุ่มตัวอย่างให้เป็นตัวแทนของประชากรที่นิยมใช้กันมี 4 วิธีคือ 1) การสุ่มแบบง่าย 2) การสุ่มแบบมีระบบ 3) การสุ่มแบบแบ่งชั้น และ 4) การสุ่มแบบแบ่งกลุ่ม
10. การสุ่มแบบง่าย เป็นวิธีสุ่มตัวอย่างที่หน่วยประชากรทุกหน่วยที่กำหนดไว้ในขอบเขตวิจัยมีโอกาสที่จะได้รับเลือกมาเป็นกลุ่มตัวอย่างเท่า ๆ กัน
11. การสุ่มแบบมีระบบ เป็นวิธีสุ่มตัวอย่างโดยการเลือกข้อมูลรายการแรกจากประชากรซึ่งถูกเรียงลำดับไว้โดยวิธีสุ่มแบบง่ายมาเพียงรายการเดียวก่อน แล้วจึงเลือกข้อมูลรายการต่อไปโดยนับเป็นช่วง ๆ ละเท่า ๆ กัน
12. การสุ่มแบบแบ่งชั้น มีหลักการว่่า ผู้ที่ทำหน้าที่สุ่มตัวอย่างจะต้องแบ่งประชากรออกเป็นชั้น ๆ เสียก่อนโดยยึดหลักว่่า ภายในแต่ละชั้นจะต้องมีลักษณะคล้ายคลึงกันมาก

สถิติอนุमान

ที่สุด และให้ระหว่างชั้นมีความแตกต่างกันมากที่สุด หลังจากนั้นจึงทำการเลือกสมาชิกภายในแต่ละชั้น ด้วยวิธีการสุ่มแบบง่ายหรือจะใช้วิธีการสุ่มแบบมีระบบภายในแต่ละชั้นก็ได้

13. การสุ่มแบบแบ่งกลุ่ม เป็นวิธีการสุ่มตัวอย่างโดยการแบ่งประชากรออกเป็นกลุ่มย่อย ๆ เสียก่อน โดยยึดหลักการว่า ในการแบ่งกลุ่มประชากรเป็นกลุ่มย่อย ๆ นี้จะต้องให้ลักษณะภายในแต่ละกลุ่มประชากรย่อยมีความแตกต่างกันมากที่สุดเท่าที่จะเป็นไปได้ ในขณะที่เดียวกันจะต้องให้มีความแตกต่างระหว่างกลุ่มน้อยที่สุดแล้วจึงใช้วิธีการสุ่มตัวอย่างแบบง่าย สุ่มตัวอย่างแต่ละกลุ่มมาอีกที

คำถามท้ายบทที่ 6

1. ให้ออกความหมายของคำต่อไปนี้
 - 1.1 สถิติอนุมาน
 - 1.2 ประชากร
 - 1.3 กลุ่มตัวอย่าง
 - 1.4 ทฤษฎีลิมิตกลาง
2. วิธีการที่ใช้ในสถิติอนุมานมีกี่วิธี จงอธิบาย
3. ค่าพารามิเตอร์ แตกต่างจากค่าสถิติอย่างไร จงอธิบาย
4. การสุ่มตัวอย่างให้เป็นตัวแทนของประชากรมีกี่วิธี และแต่ละวิธีมีหลักการสุ่มอย่างไร จงอธิบาย