



สัปดาห์ที่ 6

❖ หัวข้อการสอน

1. การวิเคราะห์ข้อมูลขั้นต้นหรือการวิเคราะห์ข้อมูลด้วยสถิติแบบบรรยาย
2. การวิเคราะห์ข้อมูลขั้นสูงหรือการวิเคราะห์ข้อมูลด้วยสถิติแบบอ้างอิง

❖ วัตถุประสงค์ เมื่อศึกษาหัวข้อนี้จบแล้ว นักศึกษาสามารถ

1. เข้าใจวิธีการวิเคราะห์ข้อมูลแต่ละวิธีโดยภาพรวม
2. เลือกใช้วิธีการวิเคราะห์ข้อมูลที่ถูกต้องเหมาะสมกับข้อมูลที่รวบรวมได้

❖ กิจกรรมการสอน

1. อาจารย์นำเข้าสู่บทเรียน
2. อาจารย์ผู้สอนบรรยายเนื้อหาพร้อมยกตัวอย่างประกอบคำบรรยาย และนำเสนอการสอนโดยใช้ Power Point
3. อาจารย์ซักถามนักศึกษาและเปิดโอกาสให้นักศึกษาซักถาม, แสดงความคิดเห็น
4. อาจารย์มอบหมายงานให้นักศึกษาตอบคำถามลงใน ใบงานที่ 6
5. อาจารย์ให้นักศึกษา อ่านเอกสารประกอบการเรียนครั้งต่อไป

❖ การวัดและการประเมินผล

1. สังเกตจากการ ถาม-ตอบของนักศึกษา
2. การตอบคำถามในใบงานที่ 6 ได้ถูกต้อง 3 ข้อ จาก 5 ข้อ
3. การสอบกลางภาค

❖ สื่อการสอน

1. เอกสารประกอบการสอน
2. คอมพิวเตอร์และเครื่องฉาย LCD
3. Power Point เรื่อง การวิเคราะห์ข้อมูล
4. ใบงานที่ 6

❖ เนื้อหา

บทที่ 4
การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเป็นการนำข้อมูลที่เก็บรวบรวมได้มาทำการจัดให้เป็นระบบระเบียบเพื่อใช้ในการอธิบายลักษณะของข้อมูล หรือเพื่อนำไปสู่การตอบคำถามของสิ่งที่สนใจอย่างรู้ โดยอาศัยเทคนิคและวิธีการทางสถิติซึ่งมีอยู่ด้วยกันมากมายหลายวิธี แต่ละวิธีก็จะมีกฎเกณฑ์และข้อจำกัดที่แตกต่างกันไป ซึ่งวิธีการวิเคราะห์ข้อมูลสามารถแบ่งได้ 2 แบบ คือ

1. การวิเคราะห์ข้อมูลขั้นต้น หรือการวิเคราะห์ข้อมูลด้วยสถิติแบบบรรยาย
2. การวิเคราะห์ข้อมูลขั้นสูง หรือการวิเคราะห์ข้อมูลด้วยสถิติแบบอ้างอิง

ในบทนี้จะกล่าวถึงการวิเคราะห์ข้อมูลทั้ง 2 แบบ อย่างคร่าว ๆ พอให้เห็นภาพกว้าง ๆ และจะกล่าวถึงการวิเคราะห์ข้อมูลด้วยสถิติแบบบรรยายอย่างละเอียดในบทต่อไป

1. การวิเคราะห์ข้อมูลด้วยสถิติแบบบรรยาย (Analysis by Descriptive Statistics)

เป็นการนำสถิติมาใช้ในการบรรยายหรืออธิบายลักษณะของข้อมูลที่เก็บรวบรวมได้ ซึ่งการบรรยายสามารถทำได้หลายวิธี โดยเป็นการวิเคราะห์ข้อมูลอย่างง่าย ๆ โดยผลจากการวิเคราะห์เบื้องต้นด้วยสถิติแบบบรรยายนั้นจะไม่มีนำไปใช้อ้างอิงกับกลุ่มอื่น แต่จะเป็นการสรุปผลเฉพาะกลุ่มที่เก็บข้อมูลมา การบรรยายลักษณะของข้อมูลด้วยสถิตินี้มีด้วยกันหลายวิธี ดังนี้

- 1.1 การแจกแจงความถี่ (Frequencies) เป็นวิธีการบรรยายลักษณะของข้อมูลด้วยการนับจำนวนค่าที่เป็นไปได้ของข้อมูลหรือตัวแปร นิยมแสดงในรูปของร้อยละด้วย
- 1.2 การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measure Central of Tendency) เป็นวิธีการบรรยายลักษณะของข้อมูลด้วยค่าตัวเลขที่เกิดจากการคำนวณหรือเลือกค่าใดค่าหนึ่งมาเป็นตัวแทนเพื่อใช้อธิบายข้อมูลที่เก็บได้ ซึ่งมีวิธีดำเนินการได้หลายวิธีดังนี้
 - 1.2.1 การคำนวณหาค่าเฉลี่ย (Mean)
 - 1.2.2 การหาค่าฐานนิยม (Mode)
 - 1.2.3 การหาค่าที่แสดงตำแหน่งของข้อมูล เช่น ค่ามัธยฐาน (Median) ควอไทล์ (Quartiles) และเปอร์เซนไทล์ (Percentiles)



- 1.3 การหาค่าการกระจายของข้อมูล (Dispersion) เป็นวิธีการหาค่าที่แสดงถึงความแตกต่างของข้อมูล หรือที่เรียกกันว่า การกระจายของข้อมูล ซึ่งจะมีวิธีการหาได้หลายวิธีดังนี้
 - 1.3.1 การหาค่าพิสัย (Range)
 - 1.3.2 การหาค่าเบี่ยงเบนมาตรฐาน (Standard Deviation)
 - 1.3.3 สัมประสิทธิ์การแปรผัน (Coefficient of Variation)
- 1.4 การหาค่าคะแนนมาตรฐาน (Standard Score) เป็นวิธีการปรับค่าของข้อมูลให้อยู่ในมาตรฐานเดียวกันเพื่อความเหมาะสมในการเปรียบเทียบข้อมูล
- 1.5 โค้งความถี่ (Frequency Curve) เป็นการอธิบายลักษณะของข้อมูลด้วยเส้นโค้งที่เกิดจากการแจกแจงความถี่

2. การวิเคราะห์ข้อมูลด้วยสถิติแบบอ้างอิง (Analysis by Inference Statistics)

เป็นการนำสถิติมาใช้ในการหาข้อสรุปเกี่ยวกับข้อมูลของประชากรโดยอาศัยค่าของข้อมูลที่ได้จากการเลือกตัวอย่างไปทำการศึกษาวิเคราะห์ โดยมีเทคนิคทางสถิติหลายเทคนิคที่จะนำมาใช้ในการหาข้อสรุป ซึ่งสามารถจำแนกเทคนิคพื้นฐานที่ควรรู้อย่างต่อไปนี้

- 2.1 การประมาณค่า (Estimation) เป็นการนำค่าของข้อมูลที่ได้จากตัวอย่าง (ค่าสถิติ) ไปประมาณค่าของข้อมูลประชากร (ค่าพารามิเตอร์) ซึ่งเป็นค่าที่มีอยู่จริงแต่ไม่สามารถหาค่ามาได้หรือถ้าหาได้ก็จะต้องใช้เวลานานและมีความยุ่งยากในการคำนวณ ในปัจจุบันมีการใช้การประมาณค่าในทุก ๆ องค์การ เช่น การประมาณยอดขายเฉลี่ยรายเดือนของบริษัทในปีหน้า เพื่อการวางแผนด้านการผลิต ด้านกำลังคน ด้านวัตถุดิบ ฯลฯ การประมาณสัดส่วนของผู้ที่ชอบทานอาหารประเภท fast food เพื่อวางแผนการจัดจำหน่ายอาหารประเภท fast food เป็นต้น การประมาณค่าแบ่งเป็น 2 ชนิด คือ
 - การประมาณค่าแบบจุด (Point Estimation) เป็นการประมาณค่าพารามิเตอร์ของประชากรด้วยตัวเลขใดตัวเลขหนึ่ง โดยใช้ค่าสถิติที่คำนวณจากข้อมูลตัวอย่าง ซึ่งเรียกค่าสถิตินี้ว่า ตัวประมาณค่าแบบจุด เช่น การประมาณรายได้เฉลี่ยต่อเดือนของคนกรุงเทพฯ ทั้งหมดด้วยรายได้เฉลี่ยต่อเดือนของคนกรุงเทพฯ ที่เป็นกลุ่มตัวอย่าง ถ้าคำนวณรายได้เฉลี่ยต่อเดือนของกลุ่มตัวอย่างคนกรุงเทพฯ ได้ 8,450 บาท ก็จะสรุปได้ว่าคนกรุงเทพฯ ทั้งหมดมีรายได้เฉลี่ยต่อเดือนประมาณ 8,450 บาท
 - การประมาณค่าแบบช่วง (Interval Estimation) เป็นการประมาณค่าพารามิเตอร์ของประชากรว่าจะอยู่ในช่วงใดช่วงหนึ่ง โดยใช้หลักการนำตัวประมาณค่าแบบจุดบวกและลบด้วยความคลาดเคลื่อนที่ยอมรับได้ (มากหรือน้อยขึ้นอยู่กับระดับ



ความเชื่อมั่นในการประมาณค่า) ซึ่งจะทำให้ได้ช่วงของการประมาณค่าที่มีทั้งค่าต่ำสุดและค่าสูงสุดของพารามิเตอร์ เช่น รายได้เฉลี่ยต่อเดือนของคนกรุงเทพฯ อยู่ในช่วง 6,500 ถึง 9,850 บาท

การจะทำการประมาณค่าแบบจุด หรือการประมาณค่าแบบช่วงนั้นขึ้นอยู่กับวัตถุประสงค์ของการใช้งาน รวมทั้งการต้องการความถูกต้องแม่นยำเพียงใด

2.2 การทดสอบสมมติฐาน (Hypothesis Testing) เป็นการนำค่าของข้อมูลที่ได้จากตัวอย่างไปทดสอบข้อสมมติหรือข้อคาดเดาเกี่ยวกับค่าพารามิเตอร์ของประชากร ซึ่งอาจจะเป็นข้อสมมติเกี่ยวกับค่าเฉลี่ย หรือค่าสัดส่วน เช่น การทดสอบสมมติฐานที่ว่าคนไทยมีอายุขัยโดยเฉลี่ยไม่เกิน 80 ปี โดยจะต้องทำการหาอายุเฉลี่ยของคนไทยตัวอย่าง แล้วนำมาทดสอบด้วยตัวสถิติทดสอบ เพื่อหาข้อสรุปว่าสมมติฐานที่กำหนดไว้เป็นจริงหรือไม่ โดยจะต้องมีการกำหนดสมมติฐานทางสถิติ ซึ่งประกอบด้วยสมมติฐานหลัก H_0 และสมมติฐานรอง H_1 พร้อมกันนี้จะต้องมีการกำหนดเกณฑ์ในการตัดสินใจ นั่นก็คือการกำหนดระดับนัยสำคัญ α (โดยส่วนมากจะใช้ $\alpha = 0.05$) ในการเลือกใช้ตัวสถิติทดสอบนั้นมีด้วยกันหลายตัว แต่ที่ใช้กันบ่อย ๆ คือ

- Z-test ใช้ในการทดสอบเกี่ยวกับ ค่าเฉลี่ย หรือค่าสัดส่วน สำหรับข้อมูลหนึ่งกลุ่มหรือสองกลุ่มที่ทราบค่าเบี่ยงเบนมาตรฐานของข้อมูลทั้งหมด ซึ่งในทางปฏิบัติแล้วการจะทราบค่านี้มีทางเป็นไปได้้น้อยมาก ทำให้ไม่ค่อยได้นำมาใช้ในการทดสอบ
- T-test ใช้ในการทดสอบเกี่ยวกับ ค่าเฉลี่ย หรือค่าสัดส่วน สำหรับข้อมูลหนึ่งกลุ่มหรือสองกลุ่มเช่นเดียวกับ Z-test โดยจะใช้แทน Z-test ในกรณีที่ไม่ทราบค่าเบี่ยงเบนมาตรฐานของข้อมูลทั้งหมด
- F-test ใช้ทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยสำหรับข้อมูลตั้งแต่ 2 กลุ่มขึ้นไป ซึ่งก็คือการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA)
- χ^2 -test (Chi-square) ใช้ในการทดสอบเกี่ยวกับความเป็นอิสระกัน ระหว่างคุณลักษณะของข้อมูลโดยใช้จำนวนหรือความถี่ของข้อมูลมาเป็นตัวทดสอบ

2.3 การหาความสัมพันธ์ (Correlation) เป็นการศึกษาแบบของความสัมพันธ์ระหว่างข้อมูลตั้งแต่ 2 ชุด หรือข้อมูลตั้งแต่ 2 คุณลักษณะขึ้นไป โดยทำการศึกษาจากข้อมูลที่ได้จากตัวอย่างเพื่อจะนำไปสู่การอธิบายหรือสรุปความสัมพันธ์ของข้อมูลประชากร เช่น การศึกษาความสัมพันธ์ระหว่างยอดขายสินค้ากับงบที่ใช้ในการโฆษณา ว่ามีความเกี่ยวข้องกันหรือไม่ การศึกษาว่าการเลือกสถานที่ท่องเที่ยวขึ้นอยู่กับอายุของ



นักท่งเกี่ยวข้องหรือไม่ เป็นต้น นอกจากจะพิจารณาว่าข้อมูลเหล่านั้นมีความสัมพันธ์กันหรือไม่แล้ว ยังสามารถหาขนาดและทิศทางการสัมพันธ์ได้อีกด้วย

2.4 การพยากรณ์ (Forecasting & Prediction) เป็นการวิเคราะห์เพื่อศึกษาหาความสัมพันธ์และรูปแบบความสัมพันธ์ของข้อมูลตัวอย่าง และนำไปสร้างตัวแบบ (Model) สำหรับใช้พยากรณ์ข้อมูลประชากรในอนาคตที่ยังไม่เกิดขึ้น โดย ในการสร้างตัวแบบให้อยู่ในรูปสมการต่าง ๆ ด้วยวิธีของสถิติอ้างอิง เช่น

- การวิเคราะห์ความถดถอย เป็นการศึกษารูปร่างความสัมพันธ์ของตัวแปร 2 กลุ่ม กลุ่มหนึ่งเรียกว่าตัวแปรอิสระ (Independent Variables) อีกกลุ่มหนึ่งเรียกว่าตัวแปรตาม (Dependent Variables) เพื่อประโยชน์ในการพยากรณ์ค่าของตัวแปรตามเมื่อทราบค่าของตัวแปรอิสระด้วยตัวแบบความสัมพันธ์ที่สร้างขึ้นจากข้อมูลตัวอย่าง
- การวิเคราะห์อนุกรมเวลา เป็นการศึกษาข้อมูลที่มีการเก็บรวบรวมแบบต่อเนื่องของช่วงเวลา โดยการสร้างความสัมพันธ์ระหว่างองค์ประกอบของข้อมูลอนุกรมเวลาซึ่งประกอบด้วยองค์ประกอบ 4 ส่วน คือ การวิเคราะห์แนวโน้ม การเปลี่ยนแปลงตามฤดูกาล การเปลี่ยนแปลงตามวัฏจักร และการเปลี่ยนแปลงเมื่อมีเหตุการณ์ผิดปกติ เพื่อประโยชน์ในการพยากรณ์ข้อมูลอนุกรมเวลาในอนาคต



❖ คำถามเพื่อทบทวน สัปดาห์ที่ 6

ใบงานที่ 6

1. จากข้อความต่อไปนี้ ข้อใดสรุปถูกต้อง
 - ก) การวิเคราะห์ข้อมูลเป็นการจัดระบบระเบียบให้กับข้อมูลที่เกิดขึ้นรวบรวมได้
 - ข) การวิเคราะห์ข้อมูลขั้นต้น เป็นการอธิบายลักษณะของข้อมูลที่เกิดขึ้นรวบรวมได้
 1. ข้อ ก. ถูก ข้อ ข. ผิด
 2. ข้อ ก. ผิด ข้อ ข. ถูก
 3. ข้อ ก. และ ข้อ ข. ผิด
 4. ข้อ ก. และ ข้อ ข. ถูก
2. ข้อใดไม่จัดเป็นการวิเคราะห์ข้อมูลด้วยสถิติแบบบรรยาย
 1. การแจกแจงความถี่
 2. การวัดค่าการกระจาย
 3. การวัดแนวโน้มเข้าสู่ส่วนกลาง
 4. การประมาณค่า

คำชี้แจง จากข้อมูลต่อไปนี้ ตอบคำถามข้อ 3-4

ในการศึกษาพฤติกรรมกรรมการอ่านหนังสือของนักศึกษามหาวิทยาลัยสยาม โดยทำการเก็บข้อมูลจากกลุ่มตัวอย่างจำนวน 396 คน พบว่ากลุ่มตัวอย่างอ่านหนังสือโดยเฉลี่ยวันละ 15 นาที

3. ข้อใดกล่าวไม่ถูกต้อง
 1. เวลาเฉลี่ยในการอ่านหนังสือของกลุ่มตัวอย่าง 15 นาทีต่อวัน จัดเป็นค่าสถิติ
 2. ประเมินได้ว่านักศึกษามหาวิทยาลัยสยามทั้งหมดอ่านหนังสือโดยเฉลี่ยวันละ 15 นาที
 3. ตามหลักการสร้างช่วงการประมาณแล้วอาจจะประมาณได้ว่าเวลาเฉลี่ยในการอ่านหนังสือของนักศึกษามหาวิทยาลัยสยามทั้งหมดอยู่ในช่วง 10 ถึง 17 นาทีต่อวัน
 4. ค่าพารามิเตอร์ที่สนใจในที่นี้คือเวลาเฉลี่ยในการอ่านหนังสือของนักศึกษามหาวิทยาลัยสยามทั้งหมด
4. ถ้ามีการคาดเดาว่านักศึกษามหาวิทยาลัยสยามทั้งหมดอ่านหนังสือโดยเฉลี่ยวันละ 12 นาที ข้อใดกล่าวไม่ถูกต้อง
 1. ต้องการทดสอบว่าค่าพารามิเตอร์เท่ากับ 12 นาทีหรือไม่
 2. ตัวสถิติทดสอบอาจจะใช้ T-test เนื่องจากไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร
 3. การกำหนดสมมติฐานทางสถิติจะต้องมี 2 ส่วนคือ H_0 และ H_1
 4. ในการทดสอบสมมติฐานไม่จำเป็นต้องใช้ข้อมูลเวลาเฉลี่ยในการอ่านหนังสือของกลุ่มตัวอย่าง



5. จากข้อความต่อไปนี้ ข้อใดสรุปถูกต้อง

- ก) การพยากรณ์ยอดขายร่วมในเดือนกรกฎาคมเมื่อมีการเก็บข้อมูลยอดขายร่วมรายเดือนตั้งแต่เดือนมกราคม ปี 2551 ถึงเดือนเมษายน ปี 2553 จะต้องใช้ความรู้ในเรื่องการวิเคราะห์การถดถอย
- ข) การศึกษาว่ายอดขายขึ้นอยู่กับราคาขายหรือไม่ จะต้องใช้ความรู้ในเรื่องของการทดสอบความสัมพันธ์
- | | |
|--------------------------|--------------------------|
| 1. ข้อ ก. ถูก ข้อ ข. ผิด | 2. ข้อ ก. ผิด ข้อ ข. ถูก |
| 3. ข้อ ก. และ ข้อ ข. ผิด | 4. ข้อ ก. และ ข้อ ข. ถูก |

★ แหล่งข้อมูลอ้างอิง

- กัลยา วานิชย์บัญชา. **หลักสถิติ**, กรุงเทพฯ : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย , 2551
- ศิริชัย พงษ์วิชัย. **สถิติเพื่อการวิจัยด้วยโปรแกรม R**, กรุงเทพฯ : บริษัท สุพีเรีย พรินติ้งเฮาส์ จำกัด , 2552.