บทที่ 3
การประยุกต์ใช้สถิติเพื่อการคาดการณ์และการแก้ปัญหา

วิธีการหนึ่งที่จะช่วยทำให้การคาดการณ์หรือการแก้ปัญหามีประสิทธิภาพมากยิ่งขึ้น คือ ใช้วิธีการทางสถิติ ซึ่งเป็นการแก้ปัญหาอย่างมีระบบ โดยเริ่มจากการเก็บรวบรวมข้อมูล เพื่อนำมาวิเคราะห์หาสาเหตุที่แท้จริงของปัญหา แล้วนำมากำหนดวิธีการแก้ไข และป้องกันปัญหาอย่างมีระบบ

การประยุกต์สถิติเพื่อใช้ในการคาดการณ์หรือการแก้ปัญหามีวิธีการมากมาย มีทั้งสถิติพรรณนาและสถิติอ้างอิง การนำสถิติไปใช้นั้น ขั้นตอนแรกต้องทราบบริบทที่เกี่ยวข้องและเป้าหมายที่ต้องการ เช่น บริบทอาจเป็นด้านการศึกษา ธุรกิจ เศรษฐศาสตร์ สังคมศาสตร์ หรือด้านอุตสาหกรรม เป้าหมายที่ต้องการอาจเป็นการบรรยายลักษณะข้อมูล หรือเป็นการประมาณค่า หรือทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย การหาความสัมพันธ์ หรือการพยากรณ์ เป็นต้น เมื่อทราบบริบทและเป้าหมายจะทำให้เลือกใช้วิธีการสถิติได้ถูกต้องมากยิ่งขึ้น

เพื่อให้มีความรู้ความเข้าใจการนำสถิติไปประยุกต์ใช้ในการคาดการณ์และการแก้ปัญหาได้อย่างมีประสิทธิภาพ จะกล่าวถึงหัวข้อตามลำดับต่อไปนี้ ข้อมูลสถิติ เทคนิคทางสถิติ โดยเนื้อหาที่กล่าวถึงจะเป็นการแนะนำให้เข้าใจมโนทัศน์ของสถิติพรรณนาและสถิติอ้างอิงบางเทคนิคที่นิยมใช้กัน พร้อมกันนั้นจะกล่าวถึงการประยุกต์ใช้สถิติเพื่อการคาดการณ์และการแก้ปัญหาในสาขาวิชาชีพหรือหน่วยงานต่าง ๆ

3.1 ข้อมูลสถิติ (Statistic data)

ข้อมูลสถิติ หมายถึง ข้อเท็จจริงของเรื่องใดเรื่องหนึ่งที่เราสนใจจะศึกษา ซึ่งอาจจะเป็นตัวเลขหรือข้อความก็ได้ เช่น รายได้ของเกษตรกรในหมู่บ้านแห่งหนึ่ง จำนวนคนไข้ที่ป่วยเป็นโรคหัวใจในโรงพยาบาลต่าง ๆ สินค้าส่งออกของประเทศไทยระหว่างปี พ.ศ. 2550 - 2555 ฯลฯ

3.2 เทคนิคทางสถิติ

คำว่า **“สถิติ”** มี 2 ความหมาย คือ

ความหมายที่ 1

สถิติ หมายถึง ข้อมูลที่เป็นตัวเลข เช่น รายได้เฉลี่ยของเกษตรกรของไทย ปริมาณข้าวโดยเฉลี่ยที่ผลิตได้ต่อไร่ ร้อยละของคนไข้ที่ป่วยเป็นโรคหัวใจ เป็นต้น ซึ่งวิธีการนำเสนอข้อมูลสถิติเหล่านี้อาจจะนำเสนอในรูปของกราฟ แผนภูมิ ตาราง หรือเป็นข้อความก็ได้

ความหมายที่ 2

สถิติ หมายถึง ความหมายที่เป็นศาสตร์ ซึ่งแบ่งได้ 2 ประเภท คือ สถิติพรรณนาและสถิติอ้างอิง

สถิติพรรณนา (Descriptive statistics) คือ สถิติที่เกี่ยวกับการจัดการ การสรุปและการนำเสนอข้อมูลในลักษณะที่ให้สาระและเป็นประโยชน์สามารถนำไปใช้ได้

สถิติอ้างอิง (Inferential statistics) คือ วิธีการทางสถิติที่ใช้ในการค้นหาลักษณะบางอย่างของประชากร โดยใช้ข้อมูลที่ได้จากกลุ่มตัวอย่าง

จากความหมายของสถิติอ้างอิง จะเห็นว่ามีคำศัพท์ ที่เข้ามาเกี่ยวข้องอย่างน้อยสองคำ คือ คำว่า “ประชากร” (Population) และ “กลุ่มตัวอย่าง” (Sample) ซึ่ง

ประชากร หมายถึง เซตของหน่วยแต่ละหน่วยที่เป็นไปได้ในกลุ่มที่สนใจศึกษา ซึ่งหน่วยแต่ละหน่วยนี้อาจเป็น บุคคล สัตว์ สิ่งของ หรือกลุ่มของค่าวัดก็ได้

กลุ่มตัวอย่าง หมายถึง บางส่วนหรือส่วนหนึ่งของประชากรที่สนใจศึกษา

เพื่อให้เข้าใจมากยิ่งขึ้นเกี่ยวกับการนำสถิติไปใช้ในการคาดการณ์หรือการแก้ปัญหา จะอธิบายเทคนิคทางสถิติที่นำไปใช้เพื่อวัตถุประสงค์เฉพาะอย่าง ดังนี้

3.2.1 สถิติพรรณนา ใช้เพื่อบรรยายลักษณะของข้อมูล สถิติที่ใช้มีดังนี้

3.2.1.1 ค่าร้อยละหรือเปอร์เซ็นต์ (Percentage) เป็นอัตราส่วนแสดงการเปรียบเทียบปริมาณใดปริมาณหนึ่งต่อ 100 เช่น

ร้อยละ 50 หรือ 50% เขียนแทนด้วย 50 : 100 หรือ 50/100

ร้อยละ 7 หรือ 7% เขียนแทนด้วย 7 : 100 หรือ 7/100

3.2.1.2 ค่าวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of central tendency) เป็นค่าที่บอกให้ทราบเกี่ยวกับภาพรวมของชุดข้อมูล ได้แก่ ค่าเฉลี่ย (Mean) มัธยฐาน (Median) และฐานนิยม (Mode) ซึ่งคำนวณได้ดังนี้

1. ค่าเฉลี่ย (Mean : $\bar{x}$) คือ ผลรวมของข้อมูลทั้งหมดหารด้วยจำนวนข้อมูล สามารถคำนวณได้ดังนี้

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

ตัวอย่างที่ 3.1 สุ่มตัวอย่างนักศึกษา 10 คน มีผลการสอบได้คะแนนดังนี้ 15, 20, 33, 45, 12, 38, 49, 29, 40, 39 จงหาค่าเฉลี่ยของข้อมูลดังกล่าว

วิธีทำ

$$\bar{x} = \frac{15+20+33+45+12+38+49+29+40+39}{10}$$

$$= \frac{320}{10} = 32$$

ดังนั้น คะแนนการสอบโดยเฉลี่ยของนักศึกษาเท่ากับ 32 คะแนน

ตัวอย่างที่ 3.2 ค่าใช้จ่ายต่อเดือนของนักศึกษา 5 คน เป็นดังนี้ 1,500, 2,000, 2,800, 1,900 และ 3,000 บาท จงหาค่าเฉลี่ยของค่าใช้จ่ายนักศึกษากลุ่มนี้

วิธีทำ

$$\bar{x} = \frac{1,500 + 2,000 + 2,800 + 1,900 + 3,000}{5}$$

$$= \frac{11,200}{5} = 2,240$$

ดังนั้น ค่าเฉลี่ยของค่าใช้จ่ายนักศึกษากลุ่มนี้ เท่ากับ 2,240 บาท

2. มัธยฐาน (Median : Me) คือ ค่าที่อยู่ในตำแหน่งกึ่งกลาง เมื่อมีการเรียงลำดับของข้อมูลจากน้อยไปมาก สามารถคำนวณตามลักษณะของข้อมูลได้ดังนี้

1. ถ้าจำนวนข้อมูล N เป็นจำนวนคี่ ค่ามัยฐาน คือ ค่าของข้อมูลในตำแหน่งที่ $\frac{N+1}{2}$

2. ถ้าจำนวนข้อมูล N เป็นจำนวนคู่ ค่ามัยฐาน คือ ให้นำค่าของข้อมูลในตำแหน่งที่ $\frac{N}{2}$ และตำแหน่งที่ $\frac{N}{2}+1$ มาเฉลี่ยกัน หรือนำมาบวกกันแล้วหารด้วย 2

ตัวอย่างที่ 3.3 จงหามัธยฐานของข้อมูลต่อไปนี้ 8, 15, 5, 10, 17, 4, 9, 13, 11

วิธีทำ เรียงลำดับข้อมูลจากค่าน้อยไปหาค่ามาก ได้ดังนี้ 4, 5, 8, 9, 10, 11, 13, 15, 17

ตำแหน่งมัธยฐาน คือ $\frac{N+1}{2} = \frac{9+1}{2} = 5$

ดังนั้น มัธยฐาน คือข้อมูลในตำแหน่งที่ 5 มีค่าเท่ากับ 10

นั่นคือ มัธยฐานของข้อมูลเท่ากับ 10

ตัวอย่างที่ 3.4 จงหามัธยฐานของข้อมูลต่อไปนี้ 20, 17, 30, 58, 44, 60, 43, 37

วิธีทำ เรียงลำดับข้อมูลจากค่าน้อยไปหาค่ามาก ได้ดังนี้ 17, 20, 30, 37, 43, 44, 58, 60

เนื่องจากจำนวนข้อมูลเป็นจำนวนคู่ ตำแหน่งมัธยฐานจึงมีสองตำแหน่ง ดังนี้

ตำแหน่งมัธยฐาน คือ $\frac{N}{2} = \frac{8}{2} = 4$

และตำแหน่งมัธยฐาน คือ $\frac{N}{2}+1 = \frac{8}{2}+1 = 4+1 = 5$

ดังนั้น มัธยฐาน คือค่าเฉลี่ยของข้อมูลในตำแหน่งที่ 4 และ 5 เท่ากับ $\frac{37+43}{2} = 40$

นั่นคือ มัธยฐานของข้อมูลเท่ากับ 40

3. ฐานนิยม (Mode : Mo) คือ ข้อมูลที่ซ้ำกันมากที่สุด ค่าฐานนิยมของข้อมูลสามารถมีได้หลาย ๆ ค่า หรือไม่มีฐานนิยมเลยในบางชุดข้อมูลก็ได้

ตัวอย่างที่ 3.5 จงหาฐานนิยมของข้อมูลต่อไปนี้ 2, 5, 4, 7, 6, 8, 8, 9, 11, 10

วิธีทำ ข้อมูลที่ซ้ำกันมากที่สุด คือ 8
ดังนั้น ฐานนิยมของข้อมูล คือ 8

ตัวอย่างที่ 3.6 จงหาฐานนิยมของข้อมูลต่อไปนี้ 10, 5, 13, 11, 8, 10, 15, 13, 23, 19

วิธีทำ ข้อมูลที่ซ้ำกันมากที่สุด คือ 10 กับ 13
ดังนั้น ฐานนิยมของข้อมูล คือ 10 กับ 13

จะเห็นว่าค่าวัดแนวโน้มเข้าสู่ส่วนกลางมี 3 ค่า การเลือกใช้ค่าใดในการบรรยายลักษณะข้อมูล จะพิจารณาได้จากตารางข้างล่างต่อไปนี้

ตารางแสดงข้อดีและข้อด้อยของค่าวัดแนวโน้มเข้าสู่ส่วนกลางทั้ง 3 ค่า

การใช้ค่าแนวโน้มเข้าสู่ส่วนกลาง	วิธีการคำนวณ	ระดับการวัดของข้อมูลที่เหมาะสม	ข้อดีและข้อด้อยของค่าวัด
ค่าเฉลี่ย	ผลรวมของข้อมูลทั้งหมดหารด้วยจำนวนข้อมูล	- อัตราส่วน - อันตรภาค	**ข้อดี** 1. บอกตำแหน่งศูนย์กลางของข้อมูล 2. ผลบวกของค่าเบี่ยงเบนจากค่าเฉลี่ยมีค่าเป็นศูนย์ **ข้อด้อย** ค่อนข้างอ่อนไหวต่อค่าสูงโด่ง
มัธยฐาน	ค่าที่อยู่ในตำแหน่งกึ่งกลางเมื่อมีการเรียงลำดับของข้อมูลจากน้อยไปมาก	- อัตราส่วน - อันตรภาค - เรียงลำดับ	**ข้อดี** ไม่ค่อยอ่อนไหวต่อค่าสูงโด่ง **ข้อด้อย** คำนวณจากค่าเพียงหนึ่งค่าหรือสองค่าเท่านั้น
ฐานนิยม	ข้อมูลที่มีความถี่สูงสุด	- อัตราส่วน - อันตรภาค - เรียงลำดับ - นามบัญญัติ	**ข้อด้อย** 1. สะท้อนลักษณะของข้อมูลไม่ชัดเจน 2. ข้อมูลบางชุดอาจจะไม่มีฐานนิยม 3. ข้อมูลบางชุดอาจจะมีฐานนิยมหลายค่า

3.2.1.3 ค่าวัดตำแหน่ง (Measures of location) เป็นค่าที่บอกให้ทราบถึงตำแหน่งของข้อมูลหนึ่ง ๆ ว่าอยู่ ณ ตำแหน่งใดของชุดข้อมูล เมื่อข้อมูลมีการเรียงลำดับจากน้อยไปหามาก ค่าวัดตำแหน่งเหล่านี้ ได้แก่ เปอร์เซ็นไทล์ (Percentile) ควอร์ไทล์ (Quartile) และเดไซล์ (Decile)

1. เปอร์เซ็นไทล์ (Percentile) เป็นการวัดตำแหน่งที่แบ่งข้อมูลทั้งหมด ที่เรียงจากน้อยไปหามาก ออกเป็น 100 ส่วนเท่า ๆ กัน แต่ละส่วนประกอบด้วยจำนวนข้อมูล $\frac{N}{100}$ จำนวน เมื่อ N แทนจำนวนข้อมูลทั้งหมด ใช้สัญลักษณ์ P_r แทน เปอร์เซ็นไทล์ที่ r

จุดที่แบ่งข้อมูลออกเป็น 100 ส่วนนั้นจะมี 99 จุด เรียกว่า เปอร์เซ็นไทล์ที่ 1 (P_1) เปอร์เซ็นไทล์ที่ 2 (P_2) เปอร์เซ็นไทล์ที่ 3 (P_3) จนกระทั่งถึง เปอร์เซ็นไทล์ที่ 99 (P_{99}) ตามลำดับ เช่น

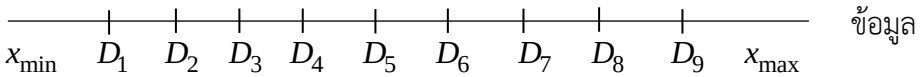
P_{60} คือ ข้อมูลที่ข้อมูลตัวอื่น ๆ ที่มีค่าน้อยหรือเท่ากับมันอยู่ 60 ส่วน และมากกว่าหรือเท่ากับมันอยู่ 40 ส่วน เมื่อแบ่งข้อมูลออกเป็น 100 ส่วน เท่า ๆ กัน

2. ควอร์ไทล์ (Quartile) เป็นการวัดตำแหน่งที่แบ่งข้อมูลทั้งหมด ที่เรียงจากน้อยไปหามาก ออกเป็น 4 ส่วนเท่า ๆ กัน แต่ละส่วนประกอบด้วยจำนวนข้อมูล $\frac{N}{4}$ จำนวน เมื่อ N แทนจำนวนข้อมูลทั้งหมด ใช้สัญลักษณ์ Q_r แทน ควอร์ไทล์ที่ r

จุดที่แบ่งข้อมูลออกเป็น 4 ส่วนนั้นจะมี 3 จุด เรียกว่า ควอร์ไทล์ที่ 1 (Q_1) ควอร์ไทล์ที่ 2 (Q_2) และควอร์ไทล์ที่ 3 (Q_3) ตามลำดับ เช่น

Q_1 คือ ข้อมูลที่ข้อมูลตัวอื่น ๆ ที่มีค่าน้อยหรือเท่ากับมันอยู่ 1 ส่วน และมากกว่าหรือเท่ากับมันอยู่ 3 ส่วน เมื่อแบ่งข้อมูลออกเป็น 4 ส่วน เท่า ๆ กัน

3. เดไซล์ (Decile) เป็นการวัดตำแหน่งที่แบ่งข้อมูลทั้งหมด ที่เรียงจากน้อยไปหามาก ออกเป็น 10 ส่วนเท่า ๆ กัน แต่ละส่วนประกอบด้วยจำนวนข้อมูล $\frac{N}{10}$ จำนวน เมื่อ N แทนจำนวนข้อมูลทั้งหมด ใช้สัญลักษณ์ D_r แทน เดไซล์ที่ r

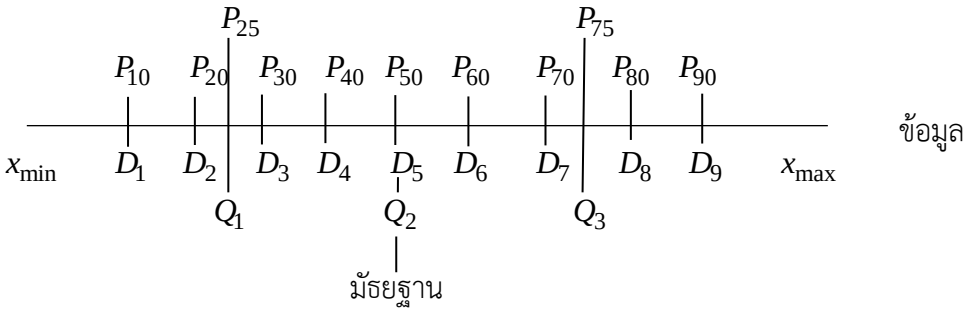


จุดที่แบ่งข้อมูลออกเป็น 10 ส่วนนั้นจะมี 9 จุด เรียกว่า เดไซล์ที่ 1 (D_1) เดไซล์ที่ 2 (D_2) เดไซล์ที่ 3 (D_3) จนกระทั่งถึง เดไซล์ที่ 9 (D_9) ตามลำดับ เช่น

D_4 คือ ข้อมูลที่ข้อมูลตัวอื่น ๆ ที่มีค่าน้อยหรือเท่ากับมันอยู่ 4 ส่วน และมากกว่าหรือเท่ากับมันอยู่ 6 ส่วน เมื่อแบ่งข้อมูลออกเป็น 10 ส่วน เท่า ๆ กัน

การเปรียบเทียบค่าของเปอร์เซ็นไทล์ ควอร์ไทล์ และเดไซล์

ข้อมูลชุดที่หนึ่งเรียงจากน้อยไปหามากและสามารถอ่านค่าของตำแหน่งด้วย P, Q, D ดังนี้



จากแผนผังข้างบน จะได้ว่า

1. $P_{10} = D_1$, $P_{20} = D_2$, $P_{30} = D_3$, … , $P_{90} = D_9$
2. $P_{25} = Q_1$, $P_{50} = Q_2$, $P_{75} = Q_3$
3. $P_{50} = Q_2 = D_5 =$ มัธยฐาน

การคำนวณหาค่าของเปอร์เซ็นไทล์ ควอร์ไทล์ และเดไซล์ ของข้อมูลที่ไม่ได้แจกแจงความถี่

ในการหาเปอร์เซ็นไทล์ ควอร์ไทล์ และเดไซล์ที่ r ของข้อมูลที่ไม่ได้แจกแจงความถี่ มีลำดับขั้นตอนการหาเหมือนกับวิธีการหามัธยฐาน ดังนี้

1. เรียงลำดับข้อมูลจากน้อยไปหามาก (หรือมากไปน้อย)
2. หาตำแหน่งของเปอร์เซ็นไทล์ ควอร์ไทล์ และเดไซล์ ที่ต้องการ โดยใช้สูตร ดังนี้

- ตำแหน่งของ $P_r = \frac{r}{100}(N+1)$
- ตำแหน่งของ $Q_r = \frac{r}{4}(N+1)$
- ตำแหน่งของ $D_r = \frac{r}{10}(N+1)$

เมื่อ r แทน ตำแหน่งของเปอร์เซ็นไทล์ ควอร์ไทล์ และเดไซล์

N แทน จำนวนข้อมูลทั้งหมด

3. ถ้าตำแหน่งที่คำนวณได้จากสูตรเป็นจำนวนเต็ม ค่าในข้อมูลซึ่งตรงกับตำแหน่งดังกล่าวจะเป็นคำตอบของเปอร์เซนไทล์ ควอร์ ไทล์ และเดไซล์ ที่ต้องการหา
4. ถ้าตำแหน่งที่คำนวณได้จากสูตรไม่เป็นจำนวนเต็ม ให้หาคำตอบของเปอร์เซนไทล์ ควอร์ไทล์ และเดไซล์ โดยการใช้เทียบบัญญัติไตรยางศ์ (ใช้สัดส่วน)

ตัวอย่างที่ 3.7 จากการทดสอบระดับสติปัญญาของนักศึกษา 9 คน ปรากฏผลดังนี้ 107, 95, 121, 103, 146, 89, 115, 135, 99 จงหา P_{40}, Q_2

วิธีทำ 1. เรียงลำดับข้อมูลจากน้อยไปหามาก ได้ดังนี้
89, 95, 99, 103, 107, 115, 121, 135, 146

2. หาตำแหน่งของ P_{40} และ Q_2

ตำแหน่งของ $P_r = \frac{r}{100}(N+1)$

$\therefore$ ตำแหน่งของ $P_{40} = \frac{40}{100}(9+1) = 4$

ตำแหน่งของ $Q_r = \frac{r}{4}(N+1)$

$\therefore$ ตำแหน่งของ $Q_2 = \frac{2}{4}(9+1) = 5$

3. หาค่าของ P_{40} และ Q_2 จะได้

ค่าของ P_{40} คือ ค่าข้อมูลซึ่งตรงกับตำแหน่ง 4 เท่ากับ 103

และค่าของ Q_2 คือ ค่าข้อมูลซึ่งตรงกับตำแหน่ง 5 เท่ากับ 107

ดังนั้น $P_{40} = 103$ และ $Q_2 = 107$

ตัวอย่างที่ 3.8 ข้อมูลน้ำหนัก (กิโลกรัม) ของนักเรียน 8 คน เป็นดังนี้ 35, 61, 43, 56 , 41, 39, 55, 54 จงหาค่าของ P_{80}, D_6

วิธีทำ 1. เรียงลำดับข้อมูลจากน้อยไปหามาก ได้ดังนี้ 35, 39, 41, 43, 54, 55, 56, 61

2. หาตำแหน่งของ P_{80} และ D_6

ตำแหน่งของ $P_r = \frac{r}{100}(N+1)$

$\therefore$ ตำแหน่งของ $P_{80} = \frac{80}{100}(8+1) = 7.2$

ตำแหน่งของ $D_r = \frac{r}{10}(N+1)$

$\therefore$ ตำแหน่งของ $D_6 = \frac{6}{10}(8+1) = 5.4$

3. หาค่าของ P_{80} และ D_6 จะได้

ค่าของ P_{80} มีค่าอยู่ระหว่าง 56 กับ 61

ตำแหน่งต่างกัน 8 - 7 = 1 ค่าข้อมูลต่างกัน 61 - 56 = 5

ถ้าตำแหน่งต่างกัน 7.2 – 7 = 0.2 ค่าข้อมูลต่างกันเท่ากับ $\frac{0.2\times5}{1} = 1$

$\therefore$ $P_{80} = 56+1 = 57$

ค่าของ D_6 มีค่าอยู่ระหว่าง 54 กับ 55

ตำแหน่งต่างกัน 6 – 5 = 1 ค่าข้อมูลต่างกัน 55 - 54 = 1

ถ้าตำแหน่งต่างกัน 5.4 – 5 = 0.4 ค่าข้อมูลต่างกันเท่ากับ $\frac{0.4\times1}{1} = 0.4$

$\therefore \quad D_6 = 54+0.4 = 54.4$

ดังนั้น $P_{80} = 57$ และ $D_6 = 54.4$

3.2.1.4 ค่าวัดการกระจาย (Measures of dispersion) เป็นค่าที่บอกให้ทราบถึงลักษณะการกระจายของข้อมูลว่าเกาะกลุ่มหรือกระจายมากอย่างไร ค่าที่ใช้วัดการกระจายที่นิยมได้แก่ พิสัย (Range) พิสัยระหว่างควอไทล์ (Interquartile range) ความแปรปรวน (Variance) และค่าเบี่ยงเบนมาตรฐาน (Standard deviation)

1. พิสัย (Range) เป็นค่าวัดการกระจายที่ไม่ซับซ้อนและเข้าใจง่าย คำนวณจากค่าเพียงสองค่า ได้แก่ ค่าสูงสุดและค่าต่ำสุดของข้อมูล สูตรที่ใช้ในการหาพิสัย คือ

พิสัย (R) = ค่าสูงสุด – ค่าต่ำสุด

ตัวอย่างที่ 3.9 จงหาพิสัยจากข้อมูลชุดนี้ 50, 32, 12, 42, 10, 29, 35, 23, 15, 30, 45

วิธีทำ พิสัย (R) = ค่าสูงสุด – ค่าต่ำสุด = 50 – 10 = 40

ดังนั้น ข้อมูลชุดนี้มีพิสัย (R) เท่ากับ 40

2. พิสัยระหว่างควอไทล์ (Interquartile range) คือ ค่าวัดการกระจายที่คำนวณจากผลต่างของข้อมูลระหว่างควอไทล์ที่ 3 และควอไทล์ที่ 1 นั่นคือ

$$IQR = Q_3 - Q_1$$

3. ความแปรปรวน (Variance) คือ ค่าเฉลี่ยของกำลังสองของผลต่างของข้อมูลกับค่าเฉลี่ย

4. ค่าเบี่ยงเบนมาตรฐาน (Standard deviation) คือ ค่ารากที่สองที่เป็นบวกของความแปรปรวน

สมการแสดงการคำนวณความแปรปรวนของ**ประชากร**

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

สมการแสดงการคำนวณค่าเบี่ยงเบนมาตรฐานของ**ประชากร**

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

โดยที่ μ คือ ค่าเฉลี่ยของประชากร
N คือ จำนวนของประชากร
σ^2 คือ ความแปรปรวนของประชากร
σ คือ ค่าเบี่ยงเบนมาตรฐานของประชากร

สมการแสดงการคำนวณความแปรปรวนของ**ตัวอย่าง**

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

สมการแสดงการคำนวณค่าเบี่ยงเบนมาตรฐานของ**ตัวอย่าง**

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

โดยที่ $\bar{x}$ คือ ค่าเฉลี่ยของตัวอย่าง
N คือ จำนวนของตัวอย่าง
S^2 คือ ความแปรปรวนของตัวอย่าง
S คือ ค่าเบี่ยงเบนมาตรฐานของตัวอย่าง

ตัวอย่างที่ 3.10 สมมติว่าข้อมูลที่เก็บรวบรวมมาเกี่ยวกับจำนวนชั่วโมงในการทำงานของพนักงาน 5 คน มีดังนี้ 15, 25, 35, 20, 30 จงหาค่าเบี่ยงเบนมาตรฐานและความแปรปรวนของข้อมูลดังกล่าว

วิธีทำ **ขั้นตอนแรก** : คำนวณค่าเฉลี่ยของประชากร

$$\mu = \frac{\sum x}{N} = \frac{15+25+35+20+30}{5} = \frac{125}{5} = 25 \quad \text{ชั่วโมง}$$

ขั้นตอนที่สอง : ให้นำเอาค่าเฉลี่ยไปลบออกจากข้อมูลแต่ละค่า ดังตารางข้างล่างนี้

x	$x-\mu$	$(x-\mu)^2$
15	15 – 25 = -10	100
25	25 – 25 = 0	0
35	35 – 25 = 10	100
20	20 – 25 = -5	25
30	30 – 25 = 5	25

ขั้นตอนที่สาม : คำนวณหาความแปรปรวน โดยแทนค่าในสูตร

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

$$= \frac{100+0+100+25+25}{5}$$

$$= \frac{250}{5} = 50$$

ดังนั้น ความแปรปรวนของจำนวนชั่วโมงในการทำงานของพนักงาน เท่ากับ 50 ชั่วโมง

หาค่าเบี่ยงเบนมาตรฐานของข้อมูลข้างต้น เราสามารถหาได้จากสูตร

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

จะได้ว่า $\sigma = \sqrt{50} = 7.07$

เนื่องจากค่าที่คำนวณได้เป็นเลขทศนิยมจึงปัดให้เป็นจำนวนเต็ม

ดังนั้น ค่าเบี่ยงเบนมาตรฐานของจำนวนชั่วโมงในการทำงานของพนักงานเท่ากับ 7 ชั่วโมง

3.2.2 สถิติอ้างอิง ใช้เพื่อการประมาณค่าและทดสอบสมมติฐานพารามิเตอร์ในกรณีที่มีการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างแทนที่จะเก็บจากประชากร สถิติที่ใช้สามารถจำแนกเป็นหมวดหมู่ตามวัตถุประสงค์หลัก ๆ ของการนำไปใช้ ดังต่อไปนี้

3.2.2.1 การเปรียบเทียบค่าเฉลี่ย เทคนิคทางสถิติที่นำมาใช้เปรียบเทียบค่าเฉลี่ย ได้แก่ t-test สำหรับการเปรียบเทียบค่าเฉลี่ยกรณีที่มีตัวอย่างกลุ่มเดียว t-test สำหรับการเปรียบเทียบค่าเฉลี่ยกรณีที่มีตัวอย่างสองกลุ่ม และการวิเคราะห์ความแปรปรวน (Analysis of variance) ใช้ในการเปรียบเทียบค่าเฉลี่ยกรณีที่มีตัวอย่างมากกว่าสองกลุ่ม

3.2.2.2 การหาความสัมพันธ์ เป็นการหาความสัมพันธ์ระหว่าง 2 ตัวแปร สถิติที่นิยมใช้ ได้แก่ สหสัมพันธ์เพียร์สัน ใช้ในกรณีที่ตัวแปรทั้งสองมีระดับการวัดอยู่ในมาตราอันตรภาค (ช่วง) และมาตราอัตราส่วน การทดสอบไคสแควร์ ใช้ในกรณีที่ตัวแปรทั้งสองมีระดับการวัดอยู่ในมาตรานามบัญญัติ และสหสัมพันธ์สเปียร์แมนลำดับ ใช้ในกรณีที่ตัวแปรทั้งสองมีระดับการวัดอยู่ในมาตราเรียงอันดับ เป็นต้น

3.2.2.3 การพยากรณ์ เทคนิคทางสถิติที่ใช้ในการพยากรณ์ ได้แก่ การวิเคราะห์การถดถอย และการวิเคราะห์อนุกรมเวลา โดยการวิเคราะห์การถดถอยจะเป็นการเก็บข้อมูลแบบภาคตัดขวาง ส่วนการวิเคราะห์อนุกรมเวลา ข้อมูลที่นำมาใช้ในวิเคราะห์จะเป็นข้อมูลอนุกรมเวลา ซึ่งก็คือ ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วง ๆ อย่างต่อเนื่องกัน โดยที่ข้อมูลของอนุกรมเวลาอาจอยู่ในลักษณะที่เป็นข้อมูลรายปี รายไตรมาส หรือรายเดือนก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์

3.3 การประยุกต์ทางสถิติ

ในการตัดสินใจเพื่อดำเนินการต่าง ๆ แน่นอนว่าจะต้องมีการใช้ข้อมูลเพื่อประเมินทางเลือกว่าควรจะเลือกทางเลือกใด การประเมินนั้นหากเรื่องที่จะต้องตัดสินใจเป็นเรื่องที่ไม่ซับซ้อน ก็จะใช้วิธีการประเมินที่ไม่ซับซ้อน นั่นคือ จากประสบการณ์อาจจะสามารถตัดสินใจหรือประเมินได้เลยว่าใช้ทางเลือกใด เช่น การเดินทางไปยังจุดหมายที่ต้องการในกรุงเทพฯ จากประสบการณ์ของเราประเมินได้ว่า ถ้าต้องการเดินทางในช่วงเวลา 07.00 – 09.00 น. จะใช้เวลานานเป็น 2 เท่าของเวลาที่ใช้ปกติ ดังนั้น เราจะต้องเผื่อเวลาการเดินทางให้มากกว่าเดิม เป็นต้น บางครั้งอาจต้องมีการตัดสินใจเรื่องที่ค่อนข้างซับซ้อนและมีปัจจัยหลายอย่างเข้ามาเกี่ยวข้อง เช่น ในทางธุรกิจ การตัดสินใจเพื่อเลือกทางเลือก จะอาศัยประสบการณ์จากผู้บริหารเพียงอย่างเดียวคงไม่ได้ มีข้อมูลมากมายที่ผู้บริหารจะต้องนำเข้ามาพิจารณาด้วย เทคนิคทางสถิติจึงเข้ามามีบทบาทในส่วนนี้ ซึ่งการเลือกใช้เทคนิคทางสถิติใดในการตัดสินใจขึ้นอยู่กับวัตถุประสงค์ของผู้ใช้ว่าต้องการอะไร ในที่นี้จะจำแนกประโยชน์ของการนำสถิติไปใช้เป็น 2 กลุ่มใหญ่ ๆ คือ การใช้สถิติพรรณนา และสถิติอ้างอิง

3.3.1 ใช้สถิติพรรณนา แทบทุกสาขาอาชีพ หน่วยงานต่าง ๆ จำเป็นอย่างยิ่งที่จะต้องมีฐานข้อมูลเบื้องต้นเพื่อใช้ในการตัดสินใจ เช่น ด้านธุรกิจ มีการคำนวณดอกเบี้ย ภาษี และหุ้น ด้านการส่งเสริมการตลาด มีการศึกษาความต้องการหรือความพึงพอใจในผลิตภัณฑ์ของลูกค้า ด้านการศึกษา สร้างฐานข้อมูลทางสถิติเพื่อการตัดสินใจ ด้านการเมืองการปกครอง เพื่อสำรวจความคิดเห็นของประชาชนด้านการเมือง และด้านสาธารณสุข เก็บข้อมูลสถิติของคนไข้ เป็นต้น

ตัวอย่างที่ 3.11 การประยุกต์สถิติพรรณนาในชีวิตประจำวัน

ในชีวิตประจำวันของคนเรา มีการนำสถิติพรรณนาไปใช้ตลอดเวลา ทั้งที่แบบรู้ตัวและแบบไม่รู้ตัว เช่น การไปซื้อของในตลาด เราจะพิจารณาเปรียบเทียบว่าควรซื้อปลาทูเป็นเข่ง ๆ หรือแยกซื้อเป็นตัว ๆ แบบใดที่เราได้ประโยชน์มากที่สุด นั่นคือ ต้องมีการคิดราคาเฉลี่ยของปลาทูต่อตัวว่าเป็นเท่าใด คุ้มหรือไม่ที่จะซื้อทั้งเข่ง เป็นต้น หรือเราอาจเลือกซื้อสินค้าหรือผลิตภัณฑ์ที่มีคนนิยมมากที่สุด ซึ่งในเรื่องนี้ก็เกี่ยวกับสถิติที่เรียกว่า “ฐานนิยม” นอกจากนี้ แม่ค้าจะต้องมีการเตรียมดอกไม้ไปขาย

มากขึ้นในวันพระเพิ่มกว่าวันปกติ เพราะจากข้อมูลการขายพบว่ามีคนอีกกลุ่มหนึ่งจะซื้อไปบูชาพระเนื่องในวันพระ

ตัวอย่างที่ 3.12 การประยุกต์สถิติพรรณนาด้านธุรกิจ

บริษัทจำหน่ายสินค้าบะหมี่สำเร็จรูป ต้องการส่งเสริมการขาย ดังนั้น จำเป็นต้องมีการสำรวจความคิดเห็นของผู้บริโภคว่ามีความคิดเห็นต่อผลิตภัณฑ์บะหมี่เป็นอย่างไร วิธีการที่บริษัทจะทำได้ ก็คือ การส่งแบบสำรวจความคิดเห็นไปถึงผู้บริโภค เมื่อได้ข้อมูลก็นำข้อมูลมาคำนวณค่าสถิติ เช่น ความถี่ ค่าเฉลี่ย หรือค่าสัมประสิทธิ์สหสัมพันธ์ เพื่อจะได้จัดโปรแกรมการส่งเสริมการขายได้อย่างมีประสิทธิภาพต่อไป

3.3.2 ใช้สถิติอ้างอิง แทบทุกสาขาและหน่วยงาน นำสถิติอ้างอิงไปใช้เพื่อการตัดสินใจไม่ว่าจะเป็นการคาดการณ์หรือการแก้ปัญหา เช่น ด้านธุรกิจ มีการพยากรณ์ยอดขาย ซึ่งอาจจะใช้การวิเคราะห์การถดถอยหรือการวิเคราะห์อนุกรมเวลา ด้านเศรษฐกิจ มีการนำเอาการวิเคราะห์การถดถอยไปใช้เพื่อพยากรณ์หรือสร้างตัวแบบสำหรับตัวแปรทางเศรษฐกิจ ด้านการศึกษา มีการนำเอาเทคนิคสถิติด้านการเปรียบเทียบค่าเฉลี่ย เช่น t-test การวิเคราะห์ความแปรปรวน เพื่อการสร้างนวัตกรรมทางการศึกษาที่มีประสิทธิภาพ เพื่อการศึกษาผลสัมฤทธิ์ทางการเรียน ด้านอุตสาหกรรม มีการนำเอาเทคนิคทางสถิติด้านการควบคุมคุณภาพสินค้าไปใช้ในกระบวนการผลิต เพื่อควบคุมคุณภาพสินค้าให้ได้มาตรฐาน นอกจากนี้ยังมีการนำเอาเทคนิคสถิติด้านการวิจัยดำเนินงาน (Operation research) เช่น การขนส่งสินค้า การควบคุมพัสดุคงคลัง การจัดแถวคอย หรือการหาทางเลือกที่ดีที่สุด ไปใช้ในการบริหารจัดการเพื่อให้เกิดประสิทธิผลสูงสุด ด้านประชากรศาสตร์ มีการนำเอาสถิติไปใช้เพื่อการพยากรณ์จำนวนประชากรในอนาคต การแพทย์และสาธารณสุข ใช้สถิติในการพยากรณ์หรือหาความน่าจะเป็นของการเกิดโรคหรือการระบาดของโรค และด้านเกษตร ใช้สถิติในการวางแผนการทดลอง เป็นต้น

ตัวอย่างที่ 3.13 การประยุกต์สถิติอ้างอิงด้านการศึกษา

สถิติอ้างอิงที่นำไปใช้ด้านการศึกษามีมากมาย เช่น การหาคุณภาพหรือการหาประสิทธิภาพของนวัตกรรมการเรียนการสอน เทคนิคทางสถิติที่นำไปใช้ เช่น การเปรียบเทียบค่าเฉลี่ยโดยใช้ t-test หรือการใช้เทคนิคการวิเคราะห์ความแปรปรวน นอกจากจะนำไปใช้ในเรื่องการหาคุณภาพหรือประสิทธิภาพของนวัตกรรมแล้ว ยังอาจนำไปใช้ในเรื่องของการประมาณจำนวนนักเรียน การวางแผนการขยายโรงเรียน เป็นต้น ซึ่งเรื่องเหล่านี้จำเป็นต้องอาศัยข้อมูลสถิติและเทคนิคทางสถิติในการวิเคราะห์ข้อมูล

ตัวอย่างที่ 3.14 การประยุกต์สถิติอ้างอิงด้านเศรษฐศาสตร์

สำนักงานสถิติแห่งชาติทำการสำรวจข้อมูลอุตสาหกรรมไอซีทีของประเทศไทย ว่าประชาชนมีความต้องการมากน้อยเพียงใด เพื่อจะรู้ว่าอุตสาหกรรมไทยด้านไอซีทีอยู่ในสภาพไหน จะได้วางแผนเพื่อการส่งเสริมได้ถูก เพราะปัจจุบันใช้การคาดเดาทั้งนั้น ที่ผ่านมาอุตสาหกรรมไอซีที มีการแบ่งตัวข้อมูลเป็น 2 ด้าน คือ ด้านดีมานด์และด้านซัพพลายก็จริง แต่มีการรวบรวมข้อมูลได้แค่

ซัพพลาย ยังไม่ได้ทำดีมานด์ว่าอยากได้อะไร ซื้ออะไร ซื้อจากใคร ข้อมูลที่สำรวจมาได้นี้วงการอุตสาหกรรมไอซีทีสามารถนำความรู้ด้านสถิติอ้างอิงมาประยุกต์ใช้ในการคำนวณดีมานด์และซัพพลายได้ เทคนิคทางสถิติที่นำไปใช้ได้ ได้แก่ การวิเคราะห์การถดถอย หรือเศรษฐมิติ เป็นต้น

กล่าวโดยสรุป การนำสถิติไปประยุกต์ใช้ในองค์กรหรือหน่วยงานต่าง ๆ สามารถนำไปใช้ตั้งแต่เรื่องง่าย ๆ เช่น การนำเสนอข้อมูลสถิติเพื่อให้อ่านหรือดูได้ง่ายขึ้น เข้าใจได้ดีขึ้น และใช้เวลาในการทำความเข้าใจได้รวดเร็วขึ้น สถิติส่วนนี้มักจะเป็นสถิติพรรณนา ซึ่งอาจจะใช้ตัวเลขไม่กี่ค่า เช่น ค่าเฉลี่ย ร้อยละ ฐานนิยม ก็จะทำให้สามารถเข้าใจภาพรวมของข้อมูลได้อย่างรวดเร็ว การนำสถิติไปใช้ในกรณีที่มีความซับซ้อน เช่น ต้องการหาความสัมพันธ์ของตัวแปรหรือการประมาณค่าตัวแปร หรือการหาสมการพยากรณ์ สถิติส่วนนี้มักจะเป็นสถิติอ้างอิง ผลที่ได้จาการวิเคราะห์จะนำไปใช้เพื่อกำหนดนโยบายหรือการตัดสินใจทางธุรกิจต่าง ๆ

แบบฝึกหัดท้ายบทที่ 3

1. จากข้อมูลต่อไปนี้ 3, 13, 5, 4, 10, 16, 20, 14, 25, 30, 28, 28, 2, 15 จงหา
    1.1 เปอร์เซ็นไทล์ที่ 80
    1.2 ควอร์ไทล์ที่ 3
    1.3 เดไซล์ที่ 6

2. กำหนดให้ข้อมูล คือ 3, 5, 6, 8, 8, 7, 10, 2 จงหาค่ามัธยฐาน

3. กำหนดให้ตัวอักษรภาษาอังกฤษ คือ W, F, A, C, C, B, D, B, F, C จงหาค่าฐานนิยมของข้อมูลชุดนี้

4. กำหนดให้ข้อมูล คือ 30, 42, 25, 34, 28, 36, 33, 44, 18 จงหาค่าต่อไปนี้
    4.1 พิสัย
    4.2 ค่าเฉลี่ย
    4.3 ค่าความแปรปรวน