สถิติเชิงพรรณนา

ความหมาย

สถิติเชิงพรรณนา หรือเรียกอีกชื่อหนึ่งว่า สถิติเชิงบรรยาย (Descriptive Statistics) เป็นวิธีการทางสถิติที่ใช้สำหรับการอธิบายตัวแปรที่เกี่ยวข้องกับกลุ่มตัวอย่างหรือประชากรเป้าหมายที่ผู้วิจัยสนใจ โดยการนำข้อมูลที่ได้จากการวัดหรือการสังเกตลักษณะหรือตัวแปรนั้นมาวิเคราะห์จัดระบบเพื่อลงสรุปให้สามารถอธิบายตัวแปรที่สนใจนั้นให้ชัดเจนขึ้น สถิติที่คำนวณได้นี้ ถ้าเป็นการคำนวณมาจากข้อมูลที่ได้จากกลุ่มตัวอย่างจะเรียกว่า *"ค่าสถิติ"* (statistics) และถ้าคำนวณจากข้อมูลที่ได้จากประชากรจะเรียกว่า "ค่าพารามิเตอร์" (parameter)

ในงานวิจัยเชิงปริมาณโดยทั่วไป มักจะคำนวณค่าสถิติ จากข้อมูลที่ได้จากกลุ่มตัวอย่างที่ศึกษาแล้วจะนำค่าสถิติที่ได้นี้ มาอ้างอิงไปยังค่าพารามิเตอร์เพื่ออธิบายประชากรโดยใช้วิธีการทางสถิติอ้างอิง (inferential statistics)

ประวัติความเป็นมา

สถิติเชิงพรรณนาเริ่มพัฒนาขึ้นมาจากงานทางด้านการปกครองโดยมุ่งที่จะเก็บข้อมูลของประชาชน เพื่อวัตถุประสงค์ในการเก็บภาษีเป็นสำคัญ โดยเริ่มจาก จอห์น เกรานท์ (John Graunt, 1620-1674) ชาวอังกฤษ ได้รวบรวมบันทึกข้อมูลสถิติการตายของประชาชนในกรุงลอนดอนระหว่างปี 1604-1661 ไว้ในหนังสือ Natural and Political Observations Made upon the Bills of Mortality และต่อมาเอ็ดมันด์ ฮาลเลย์ (Edmund Halley) นักดาราศาสตร์ สัญชาติเดียวกันก็ได้เสนอตารางวิเคราะห์ค่าใช้จ่ายรายปีของประชาชนขึ้น ซึ่งถือเป็นตารางสถิติตารางแรก

อาจกล่าวได้ว่าสถิติพรรณนาที่ใช้กันในปัจจุบันนี้ ได้เริ่มมีการพัฒนาขึ้นในราวต้นคริสต์ศตวรรษที่ 19 โดยนักสถิติ ชาวเบลเยี่ยมชื่อคิวเทลเลท (Quetelet, 1796-1984) ผู้ได้ชื่อว่า *"บิดาแห่งวิธีการเชิงปริมาณทางสังคมวิทยา"* (father of the quantitative method in socio logy) และกอลตัน (Galton, 1822-1911) เป็นผู้คิดค้นและเผยแพร่วิธีวิเคราะห์เปอร์เซ็นไทล์ มัธยฐาน สหสัมพันธ์และการกระถดถอยขึ้น และต่อมาคาร์ล เพียร์สัน (Carl Pearson, 1857-1936) นักชีววิทยาชาวอังกฤษ เป็นผู้นำสถิติดังกล่าวมาพัฒนาให้ชัดเจนยิ่งขึ้น

ระเบียบวิธีการทางสถิติเชิงพรรณนา

การใช้สถิติเชิงพรรณนาเพื่ออธิบายลักษณะหรือตัวแปรนั้น กระทำในรูปแบบต่าง ๆ กันได้หลายวิธี ขึ้นอยู่กับลักษณะของข้อมูลและจุดมุ่งหมายที่ต้องการจะอธิบาย วิธีการทางสถิติเชิงพรรณนาที่ใช้กันทั่วไป มีดังนี้

1. การแจกแจงความถี่ (frequency distribution) ซึ่งถือเป็นสถิติพื้นฐานในการอธิบายตัวแปรตัวที่ศึกษา เป็นการนำข้อมูลที่วัดได้มาแจงนับความถี่ตามค่าที่วัดได้ ว่าแต่ละค่ามีความถี่เท่าใด

จากการแจกแจงความถี่นี้ สามารถนำเสนอในรูปแผนภูมิลักษณะต่างๆ เช่น เสนอในรูปแผนภูมิแท่ง แผนภูมิภาพ หรือในรูปกราฟเส้น เพื่อให้เห็นภาพได้ชัดเจนยิ่งขึ้น

2. แนวโน้มเข้าสู่ส่วนกลาง (central tendency) การอธิบายข้อมูลในลักษณะแนวโน้มเข้าสู่ส่วนกลางเป็น การพยายามสรุปให้เห็นภาพรวมของข้อมูลทั้งชุดนั้น โดยการหาตัวแทนขึ้นมาเพียง 1 ตัวเพื่ออธิบาย ในทางปฏิบัตินิยมใช้สถิติเพื่อแสดงแนวโน้มเข้าสู่ส่วนกลางอยู่ 3 ลักษณะ คือ

2.1 ฐานนิยม (mode) เป็นสถิติที่แสดงถึงข้อมูลที่มีความถี่ของการแจกแจงสูงที่สุด

2.2 มัธยฐาน (median) เป็นสถิติที่แสดงถึงข้อมูลที่เป็นจุดกึ่งกลางของข้อมูลทั้งหมด กล่าวคือจะมีจำนวนข้อมูลที่อยู่เหนือและใต้มัธยฐานอยู่ข้างละ 50% เท่ากัน

2.3 ค่าเฉลี่ย (mean) ค่าเฉลี่ยในความหมายทั่วไป หมายถึง ค่าเฉลี่ยทางเลขคณิต (the arithmetic average) ซึ่งคำนวณได้จากการหารผลรวมของข้อมูลทุกจำนวนด้วยจำนวนข้อมูลนั้นตามสูตรดังนี้

ค่าเฉลี่ยของประชากร $= \mu = \frac{\Sigma X}{N}$

ค่าเฉลี่ยของกลุ่มตัวอย่าง $= \overline{X} = \frac{\Sigma X}{N}$

นอกจากค่าเฉลี่ยทางเลขคณิตแล้วยังมีค่าเฉลี่ยทางเรขาคณิต (geometric mean)ซึ่งคำนวณจากรากที่ N ของผลคูณของข้อมูลทั้งชุดนั้น (N : จำนวนข้อมูลในชุดนั้น) ดังนั้น

$$G = \sqrt[N]{X_1 X_2 X_3 \cdots X_N}$$

และค่าเฉลี่ยฮาร์โมนิค (harmonic mean) ซึ่งเป็นส่วนกลับของค่าเฉลี่ยเลขคณิตของส่วนกลับของข้อมูลทั้งชุดนั้น โดยคำนวณจาก

$$H = \frac{N}{\Sigma \frac{1}{X}} \quad \text{หรือ} \quad \frac{1}{H} = \frac{1}{N} \Sigma \frac{1}{X}$$

3. การกระจาย (variation or dispersion) เป็นสถิติที่แสดงให้เห็นว่าข้อมูลที่ศึกษานั้นมีความเป็นเอกพันธ์ (homogeniety) มากหรือน้อยเพียงไร ซึ่งจะทำให้อธิบายตัวแปรได้ชัดเจนและมีความหมายมากยิ่งขึ้น

สังเกตตัวอย่าง คะแนนผลการสอบของนักเรียน 2 กลุ่มดังนี้

กลุ่มที่ 1

คนที่	คะแนน
1	100
2	90
3	80
4	70
5	60

$\Sigma X = 400$

$N = 5$

$\mu = \frac{400}{5} = 80$

$Md = 80$

กลุ่มที่ 2

คนที่	คะแนน
1	82
2	81
3	80
4	79
5	78

$$\Sigma X = 400$$
$$N = 5$$
$$\mu = \frac{400}{5} = 80$$
$$Md = 80$$

จากตัวอย่าง จะเห็นว่าค่าเฉลี่ยและค่ามัธยฐานของคะแนนจากนักเรียนทั้งสองกลุ่มนี้เท่ากัน แต่แท้จริงแล้วนักเรียนสองกลุ่มนี้ก็ยังมีความสามารถที่แตกต่างกัน โดยนักเรียน ในกลุ่มที่ 1 นั้นแต่ละคนจะมีความสามารถที่แตกต่างกันสูงกว่านักเรียนในกลุ่มที่ 2 หรืออีกนัยหนึ่งนักเรียนกลุ่มที่ 2 มีความเป็นเอกพันธ์ในความสามารถสูงกว่านักเรียนกลุ่มที่ 1 ดังนั้นการหาค่าการกระจายของข้อมูล จึงทำให้สามารถอธิบายได้ชัดเจนยิ่งขึ้น

การใช้สถิติเพื่อวัดการกระจาย มี 3 ลักษณะ คือ

3.1 พิสัย (range) หมายถึง ช่วงห่างระหว่างข้อมูลที่มีค่าสูงสุดกับค่าต่ำสุด ซึ่งหาได้ด้วยการนำข้อมูลที่มีค่าต่ำสุดไปลบออกจากข้อมูลที่มีค่าสูงสุด จากตัวอย่างผลการสอบของนักเรียนสองกลุ่มข้างต้นจะได้พิสัยของคะแนนของนักเรียนกลุ่มที่ 1 จะเท่ากับ 100 - 60 = 40 คะแนน และพิสัยของคะแนนของนักเรียนกลุ่มที่ 2 เท่ากับ 82 - 78 = 4 คะแนน ซึ่งแตกต่างจากพิสัยของคะแนนของนักเรียนกลุ่ม ที่ 1 มาก

3.2 ส่วนเบี่ยงเบนควอไทล์ (quartile deviation หรือ semi-interquartile range) หมายถึง ครึ่งหนึ่งของความแตกต่างระหว่างคะแนนที่ตำแหน่งเปอร์เซ็นไทล์ที่ 75 (ตำแหน่งควอไทล์ที่ 3) กับคะแนนที่ตำแหน่งเปอร์เซ็นต์ไทล์ที่ 25 (ตำแหน่งควอไทล์ที่ 1) นั่นคือ

$$\text{ส่วนเบี่ยงเบนควอไทล์} = Q = \frac{Q_3 - Q_1}{2}$$

3.3 ความแปรปรวนและความเบี่ยงเบนมาตรฐาน (variance and standard deviation) ความแปรปรวน คือ ค่าเฉลี่ยของกำลังสองของความแตกต่างระหว่างคะแนนกับค่าเฉลี่ยของข้อมูลชุดนั้น ส่วนความเบี่ยงเบนมาตรฐานก็คำนวณ ได้จากรากที่สองของความแปรปรวน ดังนี้

$$\text{ความแปรปรวน} = \sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

$$\text{และความเบี่ยงเบนมาตรฐาน} = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

ทั้งความแปรปรวนและความเบี่ยงเบนมาตรฐานจะช่วยทำให้มองเห็นลักษณะของข้อมูลที่ต้องการอธิบายได้ชัดเจนยิ่งขึ้น ตัวอย่างเช่น ในการอธิบายความสูงและน้ำหนัก ของคนๆ หนึ่ง ว่าเขามีความสูงน้อยว่าความสูงเฉลี่ยของคนทั่วไปอยู่ 1 ความเบี่ยงเบนมาตรฐาน และน้ำหนักของเขามากกว่าน้ำหนักเฉลี่ยของคนทั่วไปอยู่ 1 ความเบี่ยงเบนมาตรฐาน เช่นนี้ทำให้พอจะมองเห็นภาพรูปร่างของคนๆ นี้ได้ค่อนข้างชัดเจนว่าเขาจะต้องเป็นคนที่อ้วนเตี้ยอย่างแน่นอน และยังบอกได้อีกว่าจะมีคนที่สูงกว่าคนๆนี้อยู่ประมาณ 84 เปอร์เซ็นต์ ในขณะเดียวกันคนๆนี้ก็มีน้ำหนักมากกว่าคนทั่วๆไปถึง 84 เปอร์เซ็นต์เช่นกัน (คือในคนทั่วๆ ไป 100 คน จะมีคนที่สูงกว่าคนๆ นี้ ประมาณ 84 คน และมีคนที่มีน้ำหนักน้อยกว่าคนๆ นี้ ประมาณ 84 คน)

4. ลักษณะโค้งปกติของข้อมูล (normality) เป็น การพรรณนาให้ทราบว่าข้อมูลมีการกระจายโค้งปกติหรือไม่ข้อมูลมีลักษณะเบ้ไปทางซ้ายหรือทางขวามีลักษณะแหลมหรือลักษณะแบน

5. ตำแหน่งสัมพันธ์ (relative position) สถิติพรรณนาในลักษณะนี้เป็นการบอกตำแหน่งของคะแนนใดคะแนนหนึ่งว่ามีความเกี่ยวข้องสัมพันธ์กับคะแนนทั้งหมดในข้อมูลชุดนั้นเช่นไร เป็นการแสดงให้เห็นว่าคะแนนตัวนั้นจะอยู่ ณ ตำแหน่งใดในการแจกแจงของข้อมูลชุดนั้น ซึ่งมีผลทำให้สามารถอธิบายคะแนนแต่ละตัวนั้นอย่างมีความหมายยิ่งขึ้น ตัวอย่าง เด็กชาย ก. ได้คะแนนจากการสอบคณิตศาสตร์ 35 คะแนน และได้คะแนนภาษาไทย 40 คะแนน จะเห็นว่าข้อมูลแค่นี้ไม่สามารถอธิบายได้ว่าเด็กชาย ก. มีความสามารถในวิชาคณิตศาสตร์และภาษาไทยอย่างไร จนกว่าจะทราบตำแหน่งของคะแนน 35 และ 40 คะแนนนี้เสีย ก่อน

สถิติที่ใช้ในการบอกตำแหน่งสัมพันธ์ได้แก่

5.1 ตำแหน่งเปอร์เซ็นไทล์ (percentile rank) เป็นการหาตำแหน่งที่ของคะแนนใด ๆ โดยเทียบค่าร้อยละของจำนวนคะแนนที่อยู่ใต้คะแนนนั้น จากตัวอย่างที่เด็กชาย ก. สอบคณิตศาสตร์ได้ 35 คะแนนนั้น เมื่อคิดตำแหน่งเปอร์เซ็นไทล์แล้ว สมมุติว่าอยู่ ณ ตำแหน่งเปอร์เซ็นไทล์ที่ 80 แสดงว่าจะมีจำนวนเด็กที่สอบคณิตศาสตร์ได้คะแนนต่ำกว่า 35 คะแนนถึงร้อยละ 80 หรืออีกนัยหนึ่งเด็กชาย ก. จะได้คะแนนคณิตศาสตร์สูงกว่าเด็กคนอื่น ๆ ถึง 80 ใน 100 คน

การคำนวณตำแหน่งเปอร์เซ็นไทล์ของคะแนนใด ๆ คำนวณได้จากการคิดหาร้อยละของความถี่สะสมที่คะแนนนั้นอยู่เหนือคะแนนอื่น

5.2 คะแนนมาตรฐาน (standard score) คะแนนมาตรฐานเป็นคะแนนแปลงรูป (derived score) ที่แสดงให้เห็นว่าคะแนนใด ๆ อยู่ห่างจากจุดอ้างอิงกี่หน่วย ซึ่งจุดอ้างอิงที่นิยมใช้กันมากก็คือการนำมาเปรียบเทียบว่าอยู่ห่างจากคะแนนเฉลี่ยกี่เท่าของความเบี่ยงเบนมาตรฐาน ซึ่งจะใช้สัญลักษณ์แทนคะแนนที่แปลงรูปนี้ว่าคะแนน Z (Z scores) ดังนั้นคะแนน Z ของคะแนนดิบใด ๆ จึงคำนวณได้จาก

$$Z = \frac{X - \mu}{\sigma}$$

ซึ่งจากคะแนนมาตรฐาน Z นี้ ก็อาจแปลงให้เป็นคะแนนมาตรฐานในรูปต่าง ๆ ได้อีก เช่น T scores, CEEB scores เป็นต้น

6. สหสัมพันธ์(relationship)เป็นสถิติที่อธิบายถึงความเข้มของความสัมพันธ์ระหว่างตัวแปรสองตัวหรือมากกว่า เรียกความเข้มของความสัมพันธ์นี้ว่า สัมประสิทธิ์สหสัมพันธ์ ซึ่งจะมีค่าระหว่าง -1.00 ถึง +1.00 ถ้าตัวแปรคู่ใดคำนวณค่าสัมประสิทธิ์สัมพันธ์ได้ใกล้ +1.00 หรือ - 1.00 แสดงว่าตัวแปรคู่นั้นมีความสัมพันธ์กันสูง โดยที่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็น + แสดงว่าตัวแปรคู่นั้นจะแปรผันตามซึ่งกันและกัน และในทางกลับกันหากค่าสัมประสิทธิ์สหสัมพันธ์เป็น - แสดงว่าตัวแปรคู่นั้นจะแปรผกผันกัน

สถิติที่ใช้หาค่าสหสัมพันธ์ที่นิยมได้แก่

6.1 สหสัมพันธ์ของอันดับ (Spearman rank order coefficient of correlation) ซึ่งเป็นสถิติที่ใช้ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอันดับที่ของตัวแปร ที่วัดในระดับเรียงอันดับ (ordinal scale) โดยคำนวณได้จากสูตร

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

6.2 สหสัมพันธ์แบบเพียร์สัน (Pearson's product - moment coefficient of correlation) เป็น การหาค่าความสัมพันธ์ระหว่างตัวแปรในกรณีที่การวัดตัวแปร อยู่ในระดับอันตรภาคชั้น (interval scale) หรืออัตราส่วน (ratio scale) ซึ่งคำนวณได้จากสูตร

$$\rho = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][(N\Sigma Y^2 - (\Sigma Y^2)]}}$$

ดิลก ดิลกานนท์

บรรณานุกรม

Gary, L.R. **Educational Research : Competencies for Analysis and Application.** 4th ed. New York : Macmillan Publishing Company, 1992.

Gravetter, Frederick J., and Larry B. Wallnau. **Statistics for Behavioral Sciences.** 3rd ed. St. Paul : West Publishing Company, 1992.

Guilford, J.P. and Benjamin Fruchter. **Fundamental Statistics in Psychology and Education.** 6th ed. Singapore : McGraw-Hill Inc. 1987.

Harris, Mary B. **Basic statistics for Behavioral Science Research.** Mass : Simon & Schuster Company, 1995.

Locther, Herman J. and McTavish, G. Donald. **Descriptive and Inferential Statistics : An Introduction.** 4th ed. Mass : Simon & Schuster Company, 1988.

Siegel, Murray R. **Theory and Problems of Statistics.** 2nd ed. New York : McGraw-Hill, Inc., 1994.

Triola, Mario F. **Elementary Statistics.** 5th ed. New York : Addison-Wesley Publishing Company Inc. 1992.