การวิเคราะห์เชิงพรรณนา และการวิเคราะห์เชิงวินิจฉัย

3

การสำรวจข้อมูลที่ได้ถูกรวบรวมไว้ นับเป็นก้าวแรกของการวิเคราะห์ข้อมูล ทั้งนี้เพื่อทำให้เกิดความเข้าใจในลักษณะของข้อมูลที่มีอยู่ โดยส่วนใหญ่มักเป็นการสังเกตลักษณะการกระจายของค่าข้อมูลในแต่ละตัวแปร ซึ่งจัดเป็นการวิเคราะห์ข้อมูลในระดับ การวิเคราะห์เชิงพรรณา รวมไปถึงการอธิบายความสัมพันธ์ระหว่างตัวแปรในชุดข้อมูล ซึ่งจัดเป็นการวิเคราะห์เชิงวินิจฉัย แม้ว่าทั้งการวิเคราะห์เชิงพรรณาและการวิเคราะห์เชิงวินิจฉัย จะเป็นการวิเคราะห์ข้อมูลเบื้องต้นที่มีวิธีการวิเคราะห์ข้อมูลที่ซับซ้อนไม่มากนัก แต่สามารถนำไปสู่คำตอบของปัญหาธุรกิจบางปัญหาได้ อีกทั้งผลลัพธ์จากการวิเคราะห์ข้อมูลในระดับขั้นนี้ยังสามารถนำไปใช้เป็นแนวทางในการตั้งสมมติฐานและการดำเนินการในการวิเคราะห์ข้อมูลในระดับขั้นที่สูงขึ้นไป ในบทนี้จะกล่าวถึงเครื่องมือทางคณิตศาสตร์และสถิติอย่างง่าย สำหรับใช้ในการวิเคราะห์เชิงพรรณาและการวิเคราะห์เชิงวินิจฉัย เพื่อเป็นแนวทางในการดำเนินการวิเคราะห์ข้อมูลและนำไปต่อยอดความรู้ของผู้อ่านต่อไปในอนาคต

3.1 สถิติศาสตร์เชิงพรรณนาด้วยตารางตัวหลัก

พิจารณาตารางข้อมูล ซึ่งประกอบด้วยคอลัมน์และแถวข้อมูล โดยแต่ละแถวข้อมูล บันทึก คุณลักษณะ ที่ สนใจ ตาม คอลัมน์ ของ ข้อมูล แต่ละ ข้อมูล (แสดงตัวอย่าง ตารางข้อมูล ดังตาราง 2.1) เมื่อนำเอาคอลัมน์ใดคอลัมน์หนึ่งของตารางข้อมูล ซึ่งหมายถึง ค่าของตัวแปรใดตัวแปรหนึ่ง ของทุกๆ ข้อมูล มาพิจารณาแล้ว เราสามารถอธิบายลักษณะการกระจายของค่าตัวแปรนั้นๆ ได้หลายวิธี วิธีหนึ่งในนั้นคือการอธิบายโดยใช้ค่าสถิติเชิงพรรณนา

ค่ากลางของข้อมูล

ค่ากลางข้อมูล เป็นค่าสถิติที่ใช้ในการบ่งบอกถึงแนวโน้มของค่าข้อมูลส่วนใหญ่ ซึ่งสามารถใช้เป็นตัวแทนค่าข้อมูลได้ เราสามารถหาค่ากลางข้อมูลได้ 3 แบบ คือ

ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) เรียกอย่างย่อว่า ค่าเฉลี่ย (Mean หรือ Average) คือ ค่าผลบวกของค่าข้อมูลทั้งหมดหารด้วยจำนวนข้อมูล กำหนดให้ $X_j = (x_{1,j}, x_{2,j}, ..., x_{n,j})$ เป็นชุดของค่าตัวแปร X_j ของทุกๆ ข้อมูลในตารางข้อมูล เราสามารถคำนวณค่าเฉลี่ยของตัวแปร X_j (แทนด้วยสัญลักษณ์ $\overline{x}_j$ เมื่อค่าตัวแปร X_j มาจากกลุ่มตัวอย่าง หรือ μ_j เมื่อค่าตัวแปร X_j มาจากประชากร) ได้จากสูตร

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j} \tag{3.1}$$

เนื่องจากการคำนวณค่าเฉลี่ยอยู่บนพื้นฐานการดำเนินการเชิงตัวเลข ดังนั้น ตัวแปรที่มีชนิดข้อมูลเป็นข้อมูลเชิงปริมาณเท่านั้น ที่จะสามารถคำนวณหาค่าเฉลี่ยได้

ตัวอย่าง 3.1

พิจารณา ชุดค่าของตัวแปรประสิทธิภาพของงาน (Job Performance: JP) $X_{\text{JP}} = (7, 10, 11, 15, 10, 10, 12, 14, 16, 12)$ จากกลุ่มตัวอย่างพนักงานบริษัท จำนวน 10 คน จะสามารถหาค่าเฉลี่ย ได้ดังนี้

$$\begin{aligned}\overline{x}_{\text{JP}} &= \frac{1}{n}\sum_{i=1}^{n} x_{i,\text{JP}} \\ &= \frac{7+10+11+15+10+10+12+14+16+12}{10} \\ &= \frac{117}{10} = 11.7\end{aligned}$$

ดังนั้น ค่าเฉลี่ยของตัวแปรประสิทธิภาพของงาน ($\overline{x}_{\text{JP}}$) มีค่าเท่ากับ 11.7

อภิธานศัพท์

ประชากร (Population)
หมายถึง กลุ่ม สมาชิก ทั้งหมด ของ สิ่งต่างๆ ที่ต้องการศึกษาหรือต้องการสรุปอ้างอิงจะเป็น คน สัตว์ สิ่งของ เหตุการณ์ ปรากฏการณ์ หรือพฤติกรรมใดๆ ก็ได้ ซึ่งอยู่ภายในขอบเขตที่กำหนด

ตัวอย่าง (Sample)
หมายถึง กลุ่ม สมาชิก ของ ประชากร ที่ถูกเลือกมาด้วยวิธีการต่างๆ เพื่อทำการศึกษาวิเคราะห์ แล้วนำผลหรือข้อสรุปที่ได้ไปใช้อ้างอิงถึงประชากร ถ้ากลุ่มสมาชิกที่เลือกมานั้นเป็นตัวแทนที่ดีของประชากร และมีจำนวนมากพอแล้วค่าที่ใช้สรุปอ้างอิงถึงประชากรจะมีความถูกต้อง หรือใกล้เคียงกับลักษณะหรือคุณสมบัติของประชากรมาก

มัธยฐาน (Median) คือ ค่าข้อมูลตรงกลางของข้อมูลทั้งหมดซึ่งถูกเรียงลำดับจากค่าน้อยไปค่ามาก หรือจากค่ามากไปค่าน้อยแล้ว การหาค่ามัธยฐานจะต้องทำการเรียงลำดับของข้อมูลก่อนเสมอ ต่อมาจึงหาตำแหน่งตรงกลาง ซึ่งสามารถแบ่งได้ 2 กรณี ดังนี้

กรณีที่ 1 จำนวนข้อมูล n เป็นเลขคี่ สามารถหาค่ามัธยฐานของตัวแปร X_j แทนด้วยสัญลักษณ์ $\tilde{x}_j$ ได้โดย

$$\tilde{x}_j = x_{\frac{(n+1)}{2},j} \tag{3.2}$$

กรณีที่ 2 จำนวนข้อมูล n เป็นเลขคู่ สามารถหาค่ามัธยฐานของตัวแปร X_j ได้โดย

$$\tilde{x}_j = \frac{x_{\frac{n}{2},j} + x_{\frac{n}{2}+1,j}}{2} \tag{3.3}$$

เนื่องจากการหาค่ามัธยฐานจะต้องทำการเรียงลำดับข้อมูล ดังนั้น ตัวแปรที่มีชนิดข้อมูลเป็นข้อมูลเชิงลำดับ หรือ ข้อมูลเชิงปริมาณ เท่านั้น ที่จะสามารถคำนวณค่ามัธยฐานได้

ตัวอย่าง 3.2

พิจารณา ชุดค่าของตัวแปรประสิทธิภาพของงาน (Job Performance: JP) $X_{\mathrm{JP}} = (7, 10, 11, 15, 10, 10, 12, 14, 16, 12)$ จากกลุ่มตัวอย่างพนักงานบริษัท จำนวน 10 คน จะสามารถหาค่ามัธยฐานได้โดยเรียงลำดับข้อมูลจากค่าน้อยไปค่ามาก ได้ดังนี้ $(7, 10, 10, 10, 11, 12, 12, 14, 15, 16)$ ชุดค่าข้อมูลนี้มีจำนวนข้อมูล n เท่ากับ 10 ซึ่งเป็นเลขคู่ ดังนั้น ค่ามัธยฐาน สามารถคำนวณได้โดยใช้สมการ (3.3) ดังนี้

$$
\begin{aligned}
\tilde{x}_{\mathrm{JP}} &= \frac{x_{\frac{n}{2},\mathrm{JP}} + x_{\frac{n}{2}+1,\mathrm{JP}}}{2} \\
&= \frac{x_{\frac{10}{2},\mathrm{JP}} + x_{\frac{10}{2}+1,\mathrm{JP}}}{2} \\
&= \frac{x_{5,\mathrm{JP}} + x_{6,\mathrm{JP}}}{2} \\
&= \frac{11 + 12}{2} \\
&= \frac{23}{2} = 11.5
\end{aligned}
$$

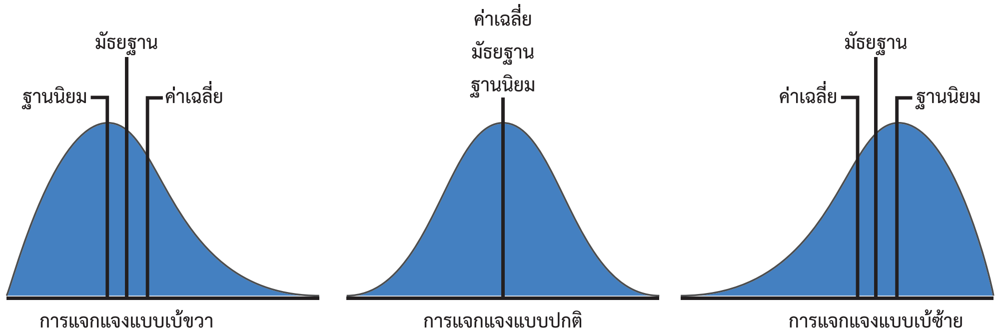
ดังนั้น ค่ามัธยฐานของตัวแปรประสิทธิภาพงาน ($\tilde{x}_{\mathrm{JP}}$) เท่ากับ 11.5

ตัวอย่าง 3.3

พิจารณา ชุดค่าของตัวแปรคะแนนสอบ TOEFL $X_{\mathrm{TOEFL}} = (118, 107, 104, 110, 109, 101, 105, 108, 110, 106, 110, 102, 104, 101, 108)$ จากกลุ่มตัวอย่างผู้เข้าสอบ TOEFL จำนวน 15 คน จะสามารถหาค่ามัธยฐานได้โดยเรียงลำดับข้อมูลจากค่าน้อยไปค่ามาก ได้ดังนี้ $(101, 101, 102, 104, 104, 105, 106, 107, 108, 108, 109, 110, 110, 110, 118)$ ชุดค่าข้อมูลนี้มีจำนวนข้อมูล n เท่ากับ 15 ซึ่งเป็นเลขคี่ ดังนั้น ค่ามัธยฐาน สามารถคำนวณได้โดยใช้สมการ (3.2) ดังนี้

$$
\begin{aligned}
\tilde{x}_{\mathrm{TOEFL}} &= x_{\frac{(n+1)}{2},\mathrm{TOEFL}} \\
&= x_{\frac{(15+1)}{2},\mathrm{TOEFL}} \\
&= x_{\frac{(16)}{2},\mathrm{TOEFL}} \\
&= x_{8,\mathrm{TOEFL}} = 107
\end{aligned}
$$

ดังนั้น ค่ามัธยฐานของตัวแปรคะแนนสอบ TOEFL ($\tilde{x}_{\mathrm{TOEFL}}$) เท่ากับ 107



ภาพ 3.1: ค่าเฉลี่ย มัธยฐาน และฐานนิยม ของลักษณะการกระจายของข้อมูลที่มีการแจกแจงแบบปกติ การแจกแจงแบบเบ้ขวา และการแจกแจงแบบเบ้ซ้าย

ฐานนิยม (Mode) คือ ค่าข้อมูลที่มีความถี่สูงที่สุด การหาค่าฐานนิยมของตัวแปร X_j แทนด้วยสัญลักษณ์ Mo_j สามารถหาได้โดยการนับความถี่ของค่าข้อมูลแต่ละค่าในชุดของค่าตัวแปร X_j และค่าข้อมูลที่มีความถี่สูงสุดจะเป็นค่าฐานนิยมของค่าตัวแปร X_j ทั้งนี้ตัวแปรที่มีชนิดข้อมูลเป็นข้อมูลนามบัญญัติ ข้อมูลเชิงลำดับ ข้อมูลระดับอัตรภาค หรือข้อมูลระดับอัตราส่วน สามารถคำนวณค่าฐานนิยมได้

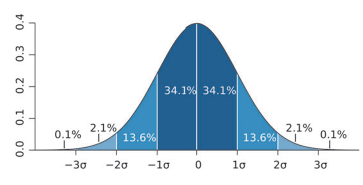
ตัวอย่าง 3.4

พิจารณา ชุดค่าของตัวแปรประสิทธิภาพของงาน (Job Performance: JP) $X_{\text{JP}} = (7, 10, 11, 15, 10, 10, 12, 14, 16, 12)$ จากกลุ่มตัวอย่างพนักงานบริษัท จำนวน 10 คน สามารถนับความถี่ของค่าข้อมูล ได้ดังนี้

ค่าข้อมูล	7	**10**	11	12	14	15	16
ความถี่	1	**3**	1	2	1	1	1

จะเห็นว่า ค่าข้อมูล 10 มีความถี่สูงที่สุด นั่นคือ ค่าความถี่เท่ากับ 3 ดังนั้น ค่าฐานนิยมของตัวแปรประสิทธิภาพของงาน (Mo_{JP}) จึงเท่ากับ 10

สำหรับชุดค่าตัวแปรหนึ่งๆ ค่ากลางของข้อมูลทั้ง 3 แบบนั้นอาจจะมีค่าเท่ากันหรือแตกต่างก็ได้ ทั้งนี้ขึ้นอยู่กับลักษณะการแจกแจงของค่าตัวแปรนั้นด้วย ดังแสดงในภาพ 3.1 ค่ากลางของข้อมูลทั้ง 3 แบบจะมีค่าเท่ากันก็ต่อเมื่อค่าข้อมูลมีการแจกแจงแบบปรกติ หากค่าข้อมูลมีลักษณะการแจกแจงแบบเบ้ซ้ายแล้ว ค่าเฉลี่ยจะมีค่าน้อยกว่าค่ามัธยฐาน และค่ามัธยฐานจะมีค่าน้อยกว่าค่าฐานนิยม ในทางกลับกันหากค่าข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวาแล้ว ค่าเฉลี่ยจะมีค่ามากกว่าค่ามัธยฐาน และค่ามัธยฐานจะมีค่ามากกว่าค่าฐานนิยม



ภาพ 3.2: แผนภาพ การ แจกแจง ปรกติ แสดง สัดส่วน พื้นที่ ใต้ โค้ง ที่ ขนาด 1 เท่า 2 เท่า และ 3 เท่า ของ ส่วน เบี่ยง เบน มาตรฐาน ของ ประชากร (σ) ทั้ง ด้าน ซ้าย และ ขวา ของ ค่า เฉลี่ย ประชากร (μ) (ที่มา: https://en.wikipedia.org/wiki/Standard_deviation)

ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน

ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD) เป็นค่าวัดทางสถิติที่บ่งบอกถึงการกระจายของค่าข้อมูลจากค่าเฉลี่ยไปทางด้านซ้ายหรือขวา ซึ่งเป็นระยะห่างเฉลี่ยกำลังสองของค่าข้อมูลทุกค่าจากค่าเฉลี่ย แสดงตำแหน่งของส่วนเบี่ยงเบนมาตราฐานบนแผนภาพการแจกแจงปรกติ ดังภาพ 3.2 ส่วนเบี่ยงเบนมาตรฐานของค่าตัวแปร X_j ของประชากร แทนด้วยสัญลักษณ์ σ_{X_j} สามารถคำนวณได้โดย

$$\sigma_{X_j} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \mu_j)^2} \tag{3.4}$$

และส่วนเบี่ยงเบนมาตรฐานของค่าตัวแปร X_j ของตัวอย่าง แทนด้วยสัญลักษณ์ s_{X_j} สามารถหาได้จากสูตร

$$s_{X_j} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2} \tag{3.5}$$

ส่วนเบี่ยงเบนมาตรฐานนั้นมีหน่วยเช่นเดียวกับหน่วยของตัวแปร การกระจายของค่าข้อมูลจากค่าเฉลี่ยแปรผันตรงกับค่าส่วนเบี่ยงเบนมาตรฐาน กล่าวคือ ส่วนเบี่ยงเบนมาตรฐานที่มีค่ามากบ่งบอกถึงค่าข้อมูลมีการกระจายตัวมาก ในทางกลับกันส่วนเบี่ยงเบนมาตรฐานมีค่าน้อยแสดงถึงการกระจายของค่าข้อมูลมีน้อยตาม

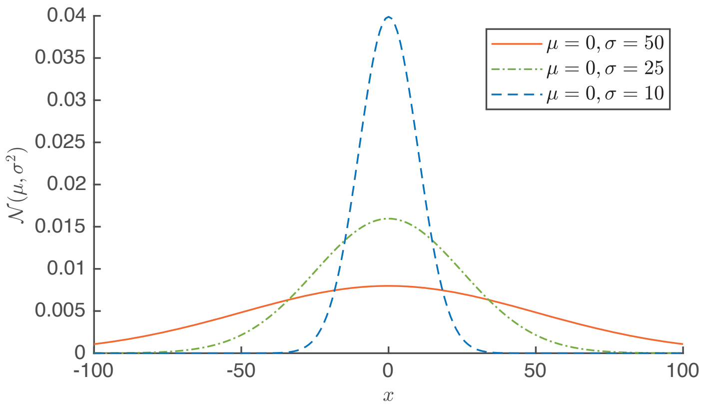
ค่าสถิติที่แสดงลักษณะการกระจายของข้อมูลที่มีความเกี่ยวเนื่องกับส่วนเบี่ยงเบนมาตรฐาน คือ **ความแปรปรวน (Variance)** ซึ่งเป็นค่ากำลังสองของส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวนของค่าตัวแปร X_j ของประชากร แทนด้วยสัญลักษณ์ $\sigma^2_{X_j}$ คำนวณได้โดย

$$\sigma^2_{X_j} = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \mu_j)^2 \tag{3.6}$$

และความแปรปรวนของค่าตัวแปร X_j ของตัวอย่าง แทนด้วยสัญลักษณ์ $s^2_{X_j}$ สามารถหาได้จากสูตร

$$s^2_{X_j} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2 \tag{3.7}$$

ค่าความแปรปรวนของข้อมูลมาก แสดงถึงการกระจายของข้อมูลมีมาก ในทางกลับกันค่าความแปรปรวนของข้อมูลน้อย บ่งบอกถึงการกระจายของข้อมูลมีน้อยด้วย ภาพ 3.3 แสดงลักษณะการกระจายของข้อมูลแบบแจกแจงปรกติ 3 ลักษณะ ซึ่งมีค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนต่างกัน ในขณะ



ภาพ 3.3: แผนภาพ การ แจกแจง ปรกติ 3 ลักษณะ ที่ มี ค่า ส่วน เบี่ยง เบน มาตรฐาน (σ) เท่ากับ 50 25 และ 10 และค่าเฉลี่ยประชากร (μ) เท่ากับ 0

ที่ค่าเฉลี่ยเลขคณิตมีค่าเท่ากัน

ตัวอย่าง 3.5

พิจารณา ชุดค่าของตัวแปรประสิทธิภาพของงาน (Job Performance: JP) $X_{\text{JP}} = (7, 10, 11, 15, 10, 10, 12, 14, 16, 12)$ จากกลุ่มตัวอย่างพนักงานบริษัท จำนวน 10 คน ซึ่งมีค่าเฉลี่ย ($\overline{x}_{\text{JP}}$) เท่ากับ 11.7 (จากตัวอย่าง 3.1) จะสามารถหาค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวน ได้ดังนี้

$x_{i,\text{JP}}$	$x_{i,\text{JP}} - \overline{x}_{\text{JP}}$	$(x_{i,\text{JP}} - \overline{x}_{\text{JP}})^2$
7	7-11.7 = -4.7	22.09
10	10-11.7 = -1.7	2.89
11	11-11.7 = -0.7	0.49
15	15-11.7 = 3.3	10.89
10	10-11.7 = -1.7	2.89
10	10-11.7 = -1.7	2.89
12	12-11.7 = 0.3	0.09
14	14-11.7 = 2.3	5.29
16	16-11.7 = 4.3	18.49
12	12-11.7 = 0.3	0.09
	รวม	66.10

จากสูตรการหาค่าความแปรปรวน

$$s^2_{X_{\text{JP}}} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,\text{JP}} - \overline{x}_{\text{JP}})^2$$

แทนค่าในสูตร จะได้

$$s^2_{X_{\mathrm{JP}}} = \frac{1}{10-1} \times 66.10$$
$$= \frac{1}{9} \times 66.10 = 7.34$$

และจากสูตรการหาค่าส่วนเบี่ยงเบนมาตรฐาน

$$s_{X_{\mathrm{JP}}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{i,\mathrm{JP}} - \overline{x}_{\mathrm{JP}})^2}$$

แทนค่าในสูตรจะได้

$$s_{X_{\mathrm{JP}}} = \sqrt{\frac{1}{10-1} \times 66.10}$$
$$= \sqrt{\frac{1}{9} \times 66.10} = \sqrt{7.34} = 2.71$$

ดังนั้น ค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปรประสิทธิภาพของงาน ($s_{X_{\mathrm{JP}}}$) เท่ากับ 2.71 และค่าความแปรปรวนของตัวแปรประสิทธิภาพของงาน ($s^2_{X_{\mathrm{JP}}}$) เท่ากับ 7.34 แสดงให้เห็นว่า ค่าข้อมูลประสิทธิภาพของงานส่วนใหญ่กระจายจากค่าเฉลี่ย ±2.71 หน่วย

ความเบ้และความโด่ง

ความเบ้ (Skewness) คือ ความไม่สมมาตรของลักษณะการแจกแจงค่าตัวแปร สามารถแบ่งออกได้ 2 ลักษณะ ได้แก่ การแจกแจงแบบเบ้ขวา (Positive Skewness) และ การแจกแจงแบบเบ้ซ้าย (Negative Skewness) แสดงลักษณะเส้นโค้งการแจกแจงแบบเบ้ขวา และเบ้ซ้าย ดังภาพ 3.1 นอกจากเราจะสามารถระบุความเบ้ของลักษณะการแจกแจงของค่าตัวแปรได้การสังเกตจากค่ากลางข้อมูลแล้ว เรายังสามารถระบุได้จากการวัดค่าสัมประสิทธิ์ความเบ้ การวัดค่าสัมประสิทธิ์ความเบ้นั้นมีวิธีการคำนวณหลายวิธี ในที่นี้จะกล่าวเฉพาะการวัดค่าสัมประสิทธิ์ความเบ้ด้วยวิธีโมเมนต์ [6] ค่าสัมประสิทธิ์ความเบ้ของค่าตัวแปร X_j แทนด้วยสัญลักษณ์ g_1 สามารถคำนวณ โดยใช้สูตรต่อไปนี้

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2\right]^{\frac{3}{2}}} \tag{3.8}$$

[6] D. N. Joanes and C. A. Gill. "Comparing measures of sample skewness and kurtosis." In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.1 (1998). url: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9884.00122

เมื่อ $m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^3$ คือ โมเมนต์ศูนย์กลาง ที่ 3 ของตัวแปร X_j และ $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2$ คือ โมเมนต์ศูนย์กลาง ที่ 2 ของตัวแปร X_j ถ้าค่าสัมประสิทธิ์ความเบ้ เท่ากับ 0 แสดงว่าค่าตัวแปรมีลักษณะการแจกแจงปกติ ถ้าค่าสัมประสิทธิ์ความเบ้ น้อยกว่า 0 แสดงว่าค่าตัวแปรมีลักษณะการแจกแจงแบบเบ้ซ้าย ในขณะที่ถ้าสัมประสิทธิ์ความเบ้มีค่ามากกว่า 0 แสดงว่าค่าตัวแปรมีลักษณะการแจกแจงแบบเบ้ขวา

ตัวอย่าง 3.6

พิจารณา ชุดค่าของตัวแปรประสิทธิภาพของงาน (Job Performance: JP) $X_{\text{JP}} = (7, 10, 11, 15, 10, 10, 12, 14, 16, 12)$ จากกลุ่มตัวอย่างพนักงานบริษัทจำนวน 10 คน ซึ่งมีค่าเฉลี่ย ($\overline{x}_{\text{JP}}$) เท่ากับ 11.7 (จากตัวอย่าง 3.1) จะสามารถหาค่าสัมประสิทธิ์ความเบ้ด้วยวิธีเมเมนต์ ได้ดังนี้

$x_{i,\text{JP}}$	$x_{i,\text{JP}} - \overline{x}_{\text{JP}}$	$(x_{i,\text{JP}} - \overline{x}_{\text{JP}})^2$	$(x_{i,\text{JP}} - \overline{x}_{\text{JP}})^3$
7	7-11.7 = -4.7	22.09	-103.82
10	10-11.7 = -1.7	2.89	-4.91
11	11-11.7 = -0.7	0.49	-0.34
15	15-11.7 = 3.3	10.89	35.94
10	10-11.7 = -1.7	2.89	-4.91
10	10-11.7 = -1.7	2.89	-4.91
12	12-11.7 = 0.3	0.09	0.03
14	14-11.7 = 2.3	5.29	12.17
16	16-11.7 = 4.3	18.49	79.51
12	12-11.7 = 0.3	0.09	0.03
	รวม	66.10	8.79

จากสูตรการหาค่าสัมประสิทธิ์ความเบ้

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2\right]^{\frac{3}{2}}}$$

แทนค่าในสูตร จะได้

$$g_1 = \frac{\frac{1}{10} \times 8.79}{\left[\frac{1}{10} \times 66.10\right]^{\frac{3}{2}}}$$
$$= \frac{0.879}{6.610^{\frac{2}{3}}}$$
$$= \frac{0.879}{3.52} = 0.25$$

ดังนั้น ค่าสัมประสิทธิ์ความเบ้ของตัวแปรประสิทธิภาพของงาน เท่ากับ 0.25 เนื่องจากค่าสัมประสิทธิ์ความเบ้มากกว่า 0 จึงกล่าวได้ว่า ค่าตัวแปรประสิทธิภาพของงานมีลักษณะการแจกแจงแบบเบ้ขวา สอดคล้องกับการพิจารณาค่ากลางข้อมูล ซึ่งมีค่าเฉลี่ยเท่ากับ 11.7 ค่ามัธยฐานเท่ากับ 11.5 และค่าฐานนิยมเท่ากับ 10 จะเห็นว่าค่าเฉลี่ยมีค่ามากกว่าค่ามัธยฐาน และค่ามัธยฐานมีค่ามากกว่าค่าฐานนิยม ซึ่งบ่งบอกถึงการแจกแจงของค่าตัวแปรประสิทธิภาพของงานมีลักษณะเบ้ขวา เช่นกัน

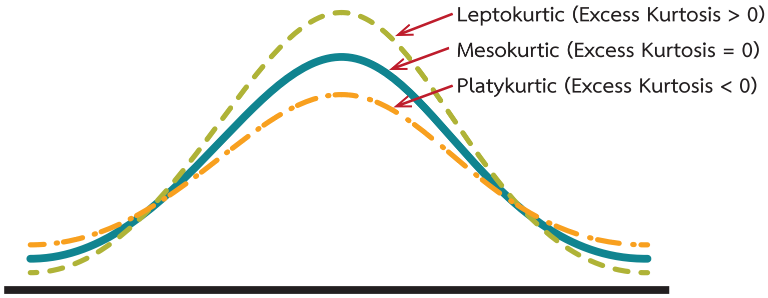
ความโด่ง (Kurtosis) เป็นค่าสถิติที่อธิบายรูปร่างของลักษณะการแจกแจงอีกค่าหนึ่ง ที่แสดงถึงความหนาของปลายหางทั้ง 2 ข้างของโค้งการแจกแจงข้อมูล [7] เราสามารถหาค่าความโด่งของตัวแปร X_j แทนด้วยสัญลักษณ์ g_2 ด้วยวิธีโมเมนต์ ได้จากสูตรต่อไปนี้

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2\right]^2} - 3 \tag{3.9}$$

เมื่อ $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^4$ คือ โมเมนต์ศูนย์กลาง ที่ 4 ของตัวแปร X_j และ $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2$ คือ โมเมนต์ศูนย์กลาง ที่ 2 ของตัวแปร X_j ค่าความโด่งที่คำนวณได้จากสมการ (3.9) เรียกว่า ค่าอิคเซส-เคอร์โทซิส (Excess Kurtosis) ลักษณะความโด่งของโค้งการแจกแจงแบ่งออกได้ 3 ลักษณะ ตามค่าอิคเซส-เคอร์โทซิส (แสดงดังภาพ 3.4) ดังนี้

- พลาติเคอร์ติค (Platykurtic) เป็นการแจกแจงที่มีความโด่งต่ำกว่าความโด่งของการแจกแจงปกติ ซึ่งจะมีค่าอิคเซส-เคอร์โทซิส น้อยกว่า 0
- เมโซเคอร์ติค (Mesokurtic) เป็นการแจกแจงที่มีความโด่งเท่ากับความโด่งของการแจกแจงปกติ ซึ่งจะมีค่าค่าอิคเซส-เคอร์โทซิส เท่ากับ 0
- เลปโตเคอร์ติค (Leptokurtic) เป็นการแจกแจงที่มีความโด่งมากกว่าความโด่งของการแจกแจงปกติ ซึ่งจะมีค่าค่าอิคเซส-เคอร์โทซิส มากกว่า 0

[7] Peter H. Westfall. “Kurtosis as Peakedness, 1905–2014. R.I.P.” In: ***The American Statistician*** 68.3 (2014). url: https://doi.org/10.1080/00031305.2014.917055



ภาพ 3.4: แผนภาพการแจกแจง 3 ลักษณะตามค่า ค่าอิคเซส-เคอร์โทซิส

ตัวอย่าง 3.7

พิจารณา ชุดค่าของตัวแปรประสิทธิภาพของงาน (Job Performance: JP) $X_{\text{JP}} = (7, 10, 11, 15, 10, 10, 12, 14, 16, 12)$ จากกลุ่มตัวอย่างพนักงานบริษัท จำนวน 10 คน ซึ่งมีค่าเฉลี่ย ($\overline{x}_{\text{JP}}$) เท่ากับ 11.7 (จากตัวอย่าง 3.1) จะสามารถหาค่าอิคเซส-เคอร์โทซิส ได้ดังนี้

$x_{i,\text{JP}}$	$x_{i,\text{JP}} - \overline{x}_{\text{JP}}$	$(x_{i,\text{JP}} - \overline{x}_{\text{JP}})^2$	$(x_{i,\text{JP}} - \overline{x}_{\text{JP}})^4$
7	7-11.7 = -4.7	22.09	487.97
10	10-11.7 = -1.7	2.89	8.35
11	11-11.7 = -0.7	0.49	0.24
15	15-11.7 = 3.3	10.89	118.59
10	10-11.7 = -1.7	2.89	8.35
10	10-11.7 = -1.7	2.89	8.35
12	12-11.7 = 0.3	0.09	0.01
14	14-11.7 = 2.3	5.29	27.98
16	16-11.7 = 4.3	18.49	341.88
12	12-11.7 = 0.3	0.09	0.01
	รวม	66.10	1001.73

จากสูตรการหาค่าสัมประสิทธิ์ความเบ้

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)^2\right]^2} - 3$$

แทนค่าในสูตร จะได้

$$
\begin{aligned}
g_2 &= \frac{\frac{1}{10} \times 1001.73}{\left[\frac{1}{10} \times 66.10\right]^2} - 3 \\
&= \frac{100.173}{6.610^2} - 3 \\
&= \frac{100.173}{43.69} - 3 \\
&= 2.29 - 3 = -0.71
\end{aligned}
$$

ดังนั้น ค่าอิคเซส-เคอร์โทซิสของตัวแปรประสิทธิภาพของงาน เท่ากับ -0.71 จากค่าอิคเซส-เคอร์โทซิสมีค่าน้อยกว่า 0 จึงกล่าวได้ว่า โค้งการแจกแจงของค่าตัวแปรประสิทธิภาพของงานมีความโด่งน้อยกว่าการแจกแจงปกติ และมีลักษณะความโด่งแบบพลาติเคอร์ติด

ความแปรปรวนร่วม

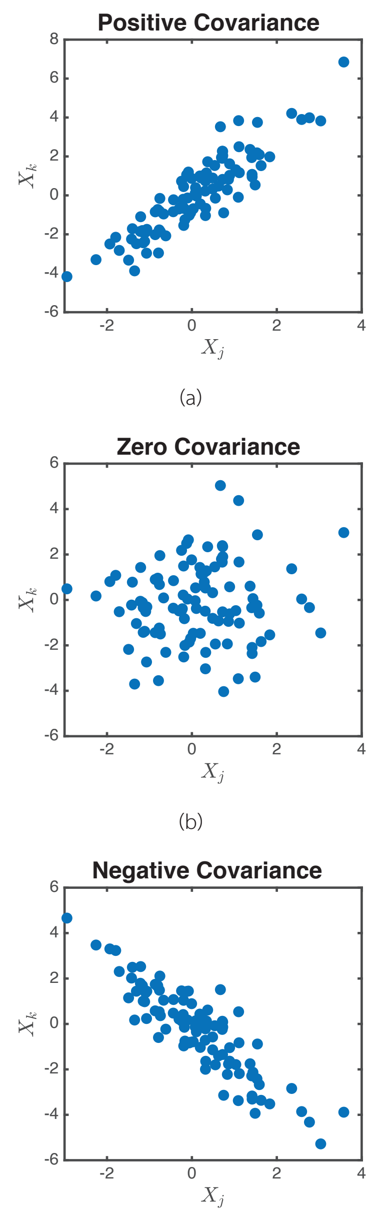
ความแปรปรวนร่วม (Covariance) เป็นค่าชี้วัดการเปลี่ยนแปลงของค่าตัวแปรตัวใดตัวหนึ่งเมื่อมีการเปลี่ยนแปลงในค่าตัวแปรอีกตัวหนึ่ง ซึ่งแสดงระดับความสัมพันธ์เชิงเส้นตรงของตัวแปร 2 ตัวแปร กำหนดให้ $X_j = (x_{1,j}, x_{2,j}, ..., x_{n,j})$ และ $X_k = (x_{1,k}, x_{2,k}, ..., x_{n,k})$ เป็นชุดของค่าตัวแปร X_j และ X_k ของข้อมูลจำนวน n ข้อมูล โดยค่า $x_{i,j}$ และ $x_{i,k}$ เป็นค่าตัวแปร X_j และ X_k ของข้อมูลที่ i ค่าความแปรปรวนร่วมระหว่างตัวแปร X_j และ X_k ของประชากร แทนด้วยสัญลักษณ์ σ_{X_j,X_k} สามารถคำนวณได้จากสูตร

$$\sigma_{X_j,X_k} = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \mu_j)(x_{i,k} - \mu_k) \tag{3.10}$$

และค่าความแปรปรวนร่วมระหว่างตัวแปร X_j และ X_k ของตัวอย่าง แทนด้วยสัญลักษณ์ s_{X_j,X_k} สามารถคำนวณได้จากสูตร

$$s_{X_j,X_k} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,j} - \overline{x}_j)(x_{i,k} - \overline{x}_k) \tag{3.11}$$

ค่าความแปรปรวนร่วม มีค่าที่เป็นไปได้อยู่ระหว่าง $-\infty$ และ ∞ ค่าความแปรปรวนร่วมที่มีค่ามากกว่า 0 แสดงถึงความสัมพันธ์ระหว่างตัวแปรทั้งสอง โดยเมื่อค่าของตัวแปรหนึ่งมีค่ามาก ค่าของอีกตัวแปรหนึ่งมักมีค่ามากด้วย ในทางเดียวกันเมื่อค่าตัวแปรหนึ่งมีค่าน้อย ค่าของอีกตัวแปรหนึ่งมักมีค่าน้อยตามไปด้วย ในกรณีที่ค่าความแปรปรวนร่วมมีค่าน้อยกว่า 0 แสดงถึงความ



(c)

ภาพ 3.5: แผนภาพ การกระจาย ของ ค่าข้อมูลที่สัมพันธ์กับค่าความแปรปรวนร่วม 3 ลักษณะ คือ (a) ค่าความแปรปรวนร่วมมีค่ามากกว่า 0 (b) ค่าความแปรปรวนร่วมมีค่าเท่ากับ 0 (c) ค่าความแปรปรวนร่วมมีค่าน้อยกว่า 0

สัมพันธ์ระหว่างสองตัวแปรในทางตรงกันข้าม กล่าวคือเมื่อค่าของตัวแปรหนึ่งมีค่ามาก ค่าของอีกตัวแปรหนึ่งมักมีค่าน้อย ในทางกลับกันเมื่อค่าตัวแปรหนึ่งมีค่าน้อย ค่าของอีกตัวแปรหนึ่งมักมีค่ามาก ส่วนค่าความแปรปรวนร่วมที่มีค่าเท่ากับ 0 บ่งบอกถึงค่าของตัวแปรทั้งสองตัวแปรไม่มีความสัมพันธ์เชิงเส้นตรงต่อกัน แสดงแผนภาพการกระจายของค่าข้อมูลระหว่าง 2 ตัวแปร และค่าความแปรปรวนร่วมที่สัมพันธ์กัน ดังภาพ 3.5

เนื่องจากค่าความแปรปรวนร่วมสามารถแสดงระดับความสัมพันธ์ระหว่างตัวแปรได้เพียง 2 ตัวแปร เราจึงไม่สามารถหาค่าความแปรปรวนร่วมเพียงค่าเดียว ในกรณีที่จำนวนตัวแปรที่สนใจมีมากกว่า 2 ตัวแปรได้ อย่างไรก็ตาม เราสามารถพิจารณาค่าความแปรปรวนร่วมของแต่ละคู่ของตัวแปรที่ทำการศึกษา โดยสร้างเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) ขนาด $d \times d$ เมื่อ d คือ จำนวนตัวแปรที่ศึกษา กำหนด $\mathbf{X} = (X_1, X_2, ..., X_d)$ เป็นชุดตัวแปรบนชุดข้อมูล $\mathbf{D}$ เมทริกซ์ความแปรปรวนร่วมของชุดตัวแปร $\mathbf{X}$ แทนด้วยสัญลักษณ์ $\mathrm{cov}(\mathbf{X})$ สามารถสร้างได้ ดังนี้

$$\mathrm{cov}(\mathbf{X}) = \begin{bmatrix} \sigma_{X_1,X_1} & \sigma_{X_1,X_2} & \cdots & \sigma_{X_1,X_d} \\ \sigma_{X_2,X_1} & \sigma_{X_2,X_2} & \cdots & \sigma_{X_2,X_d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_d,X_1} & \sigma_{X_d,X_2} & \cdots & \sigma_{X_d,X_d} \end{bmatrix} \tag{3.12}$$

จากคุณสมบัติของค่าความแปรปรวนร่วม $\sigma_{X_j,X_k} = \sigma_{X_k,X_j}$ จึงทำให้เมทริกซ์ความแปรปรวนร่วมเป็นเมทริกซ์สมมาตร (Symmetric Matrix) ซึ่งมีค่าในตำแหน่ง (j,k) เท่ากับค่าในตำแหน่ง (k,j) และจากคุณสมบัติ $\sigma_{X_j,X_j} = \sigma^2_{X_j}$ ทำให้ค่าในเส้นทแยงมุม (ตำแหน่ง (j,k) ที่ $j = k$) ของเมทริกซ์ แสดงค่าความแปรปรวนของค่าตัวแปร X_j นั่นเอง

ตัวอย่าง 3.8

พิจารณา ชุดข้อมูลจากการสำรวจระดับเชาว์ปัญญา (IQ) และประสิทธิภาพของงาน (Job Performance: JP) จากตัวอย่างพนักงานบริษัท จำนวน 10 คน (แสดงดังตาราง ด้านล่าง) โดยค่าเฉลี่ยระดับเชาว์ปัญญา ($\bar{x}_{\mathrm{IQ}}$) เท่ากับ 111.5 และค่าเฉลี่ยประสิทธิภาพของงาน ($\bar{x}_{\mathrm{JP}}$) เท่ากับ 11.7 จะสามารถหาค่าความแปรปรวนร่วมระหว่างตัวแปรระดับเชาว์ปัญญาและประสิทธิภาพของงาน ได้ดังนี้

$x_{i,\text{IQ}}$	$x_{i,\text{JP}}$	$x_{i,\text{IQ}} - \overline{x}_{\text{IQ}}$	$x_{i,\text{JP}} - \overline{x}_{\text{JP}}$	$(x_{i,\text{IQ}} - \overline{x}_{\text{IQ}})(x_{i,\text{JP}} - \overline{x}_{\text{JP}})$
99	7	-12.5	-4.7	58.75
105	10	-6.5	-1.7	11.05
105	11	-6.5	-0.7	4.55
106	15	-5.5	3.3	18.15
108	10	-3.5	-1.7	5.95
112	10	0.5	-1.7	-0.85
113	12	1.5	0.3	0.45
115	14	3.5	2.3	8.05
118	16	6.5	4.3	27.95
134	12	22.5	0.3	6.75
			รวม	140.80

จากสูตรการหาค่าความแปรปรวนร่วมของตัวอย่าง

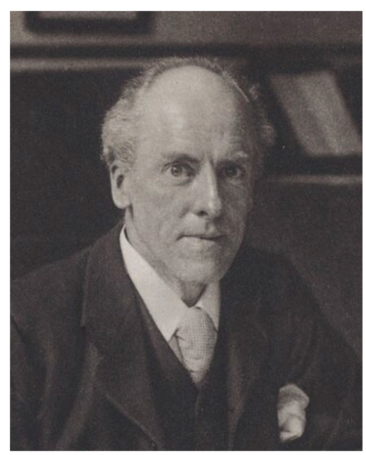
$$s_{X_{\text{IQ}},X_{\text{JP}}} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,\text{IQ}} - \overline{x}_{\text{IQ}})(x_{i,\text{JP}} - \overline{x}_{\text{JP}})$$

แทนค่าในสูตร จะได้

$$\begin{aligned} s_{X_{\text{IQ}},X_{\text{JP}}} &= \frac{1}{10-1}\sum_{i=1}^{10}(x_{i,\text{IQ}} - \overline{x}_{\text{IQ}})(x_{i,\text{JP}} - \overline{x}_{\text{JP}}) \\ &= \frac{1}{9} \times 140.80 \\ &= 15.64 \end{aligned}$$

ดังนั้น ค่าความแปรปรวนร่วมระหว่างตัวแปรระดับเชาว์ปัญญาและประสิทธิภาพของงาน เท่ากับ 15.64 บ่งชี้ว่าค่าตัวแปรระดับเชาว์ปัญญาและประสิทธิภาพของงานมีความสัมพันธ์กัน โดยเมื่อค่าตัวแปรระดับเชาว์ปัญญามีค่ามาก ค่าตัวแปรประสิทธิภาพของงานจะมีค่ามากด้วย ในขณะเดียวกันเมื่อค่าตัวแปรระดับเชาว์ปัญญามีค่าน้อย ค่าตัวแปรประสิทธิภาพของงานจะมีค่าน้อยด้วย

อย่างไรก็ตามขนาดของค่าความแปรปรวนร่วมนั้นยากที่นำมาแปลความหมายและเปรียบเทียบกัน เนื่องจากค่าความแปรปรวนร่วมไม่ได้ถูกทำให้เป็นบรรทัดฐาน อีกทั้งยังขึ้นอยู่กับขนาดของค่าตัวแปร ค่าวัดระดับความสัมพันธ์เชิงเส้นตรงของตัวแปร 2 ตัวแปร อีกค่าหนึ่ง คือ **ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Correlation Coefficient)** ซึ่งเป็นค่าความแปรปรวนร่วมระหว่างตัวแปร 2 ตัวแปรที่ถูกทำให้เป็นบรรทัดฐาน โดยการหารด้วยผลคูณของค่า



ภาพ 3.6: คาร์ล เพียร์สัน (Karl Pearson: ค.ศ. 1857-1936) นักคณิตศาสตร์ และนักชีวสถิติชาวอังกฤษ ผู้พัฒนาสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (ที่มา: https://en.wikipedia.org/wiki/Karl_Pearson)

ส่วนเบี่ยงเบนมาตรฐานของทั้งสองตัวแปร ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างตัวแปร X_j และ X_k สำหรับประชากร แทนด้วยสัญลักษณ์ ρ_{X_j,X_k} สามารถคำนวณได้จาก

$$\rho_{X_j,X_k} = \frac{\sigma_{X_j,X_k}}{\sigma_{X_j}\sigma_{X_k}} \tag{3.13}$$

และ ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างตัวแปร X_j และ X_k สำหรับตัวอย่าง แทนด้วยสัญลักษณ์ r_{X_j,X_k} สามารถคำนวณได้จาก

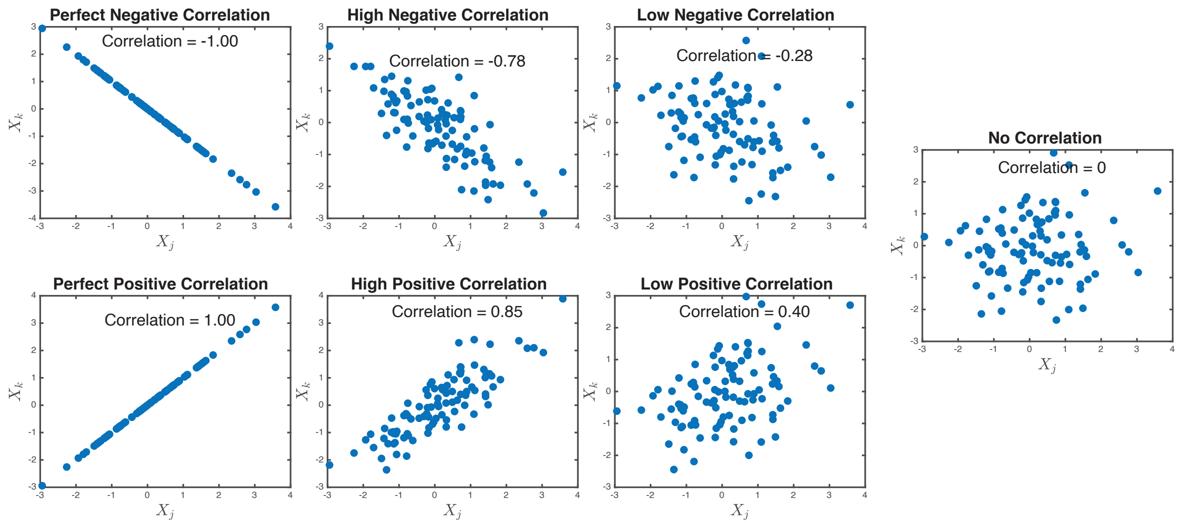
$$r_{X_j,X_k} = \frac{s_{X_j,X_k}}{s_{X_j}s_{X_k}} \tag{3.14}$$

ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน มีค่าที่เป็นไปได้อยู่ระหว่าง -1 และ 1 ค่าสัมประสิทธิ์สหสัมพันธ์ ที่มีค่าเท่ากับ 1 บ่งบอกถึงความสัมพันธ์เชิงเส้นตรงอย่างสมบูรณ์แบบระหว่างค่าตัวแปรทั้งสอง เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้นค่าของอีกตัวแปรหนึ่งจะเพิ่มขึ้นตามไปด้วย ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเท่ากับ -1 แสดงถึงความสัมพันธ์เชิงเส้นตรงในลักษณะตรงข้ามอย่างสมบูรณ์ของตัวแปรทั้งสอง โดยเมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้นค่าของอีกตัวแปรหนึ่งจะลดลง ในกรณีที่ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ 0 แสดงถึงค่าของตัวแปรทั้งสองตัวแปรไม่มีความสัมพันธ์เชิงเส้นตรงต่อกัน ค่าสัมบูรณ์ (Absolute Value) ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ยังชี้วัดระดับความสัมพันธ์ของค่าตัวแปรทั้งสองด้วย กล่าวคือ ค่าสัมบูรณ์ของค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่ามาก บ่งบอกถึงค่าตัวแปรทั้งสองมีความสัมพันธ์กันอย่างมากกว่า ในขณะที่ค่าสัมบูรณ์ของค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าน้อย แสดงถึงความสัมพันธ์ในระดับต่ำระหว่างตัวแปรทั้งสองด้วย แสดงแผนภาพการกระจายของค่าข้อมูลระหว่าง 2 ตัวแปร และสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่สัมพันธ์กัน ดังภาพ 3.7

เช่นเดียวกับค่าความแปรปรวนร่วม เราสามารถอธิบายความสัมพันธ์สำหรับชุดข้อมูลที่มีจำนวนตัวแปรมากกว่า 2 ตัวแปรได้ โดยการสร้างเมทริกซ์สหสัมพันธ์เพียร์สัน (Pearson Correlation Matrix) สำหรับอธิบายความสัมพันธ์ระหว่างแต่ละคู่ของตัวแปร เมทริกซ์สหสัมพันธ์เพียร์สัน ของชุดตัวแปร X แทนด้วยสัญลักษณ์ corr(X) สามารถสร้างได้ ดังนี้

$$\mathrm{corr}(\mathbf{X}) = \begin{bmatrix} \rho_{X_1,X_1} & \rho_{X_1,X_2} & \cdots & \rho_{X_1,X_d} \\ \rho_{X_2,X_1} & \rho_{X_2,X_2} & \cdots & \rho_{X_2,X_d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{X_d,X_1} & \rho_{X_d,X_2} & \cdots & \rho_{X_d,X_d} \end{bmatrix} \tag{3.15}$$

เมทริกซ์สหสัมพันธ์เพียร์สันเป็นเมทริกซ์สมมาตร (Symmetric Matrix) ซึ่งมีค่าในตำแหน่ง (j,k) เท่ากับค่าในตำแหน่ง (k,j) และค่าในเส้นทแยงมุมของ



ภาพ 3.7: แผนภาพการกระจายของค่าข้อมูลที่สัมพันธ์กับค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน 7 ลักษณะ

เมทริกซ์ ซึ่งแสดงค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันภายในตัวแปรเดียวกัน มีค่าเท่ากับ 1 เสมอ

ตัวอย่าง 3.9

พิจารณา ชุดข้อมูลจากการสำรวจระดับเชาว์ปัญญา (IQ) และประสิทธิภาพของงาน (Job Performance: JP) จากตัวอย่างพนักงานบริษัท จำนวน 10 คน โดยค่าส่วนเบี่ยงเบนมาตรฐานของค่าตัวแปรระดับเชาว์ปัญญา (s_{IQ}) เท่ากับ 9.70 ค่าส่วนเบี่ยงเบนมาตรฐานของค่าตัวแปรประสิทธิภาพของงาน (s_{JP}) เท่ากับ 2.71 และค่าความแปรปรวนร่วมระหว่างตัวแปรระดับเชาว์ปัญญาและประสิทธิภาพของงาน เท่ากับ 15.64 (จากตัวอย่าง 3.8) จะสามารถหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างตัวแปรระดับเชาว์ปัญญาและประสิทธิภาพของงาน ได้ดังนี้

จากสูตรการหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันของตัวอย่าง

$$r_{X_{\text{IQ}},X_{\text{JP}}} = \frac{s_{\text{IQ},X_{\text{JP}}}}{s_{\text{IQ}}s_{\text{JP}}} \tag{3.16}$$

แทนค่าในสูตร จะได้

$$\begin{aligned} r_{X_{\text{IQ}},X_{\text{JP}}} &= \frac{15.64}{9.70 \times 2.71} \\ &= \frac{15.64}{26.287} \\ &= 0.59 \end{aligned}$$

ดังนั้น ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างตัวแปรระดับเชาว์

ปัญญาและประสิทธิภาพของงาน เท่ากับ 0.59 บ่งชี้ว่าค่าตัวแปรระดับเชาว์ปัญญา และประสิทธิภาพของงานมีความสัมพันธ์เชิงเส้นตรงต่อกัน โดยเมื่อค่าตัวแปรระดับเชาว์ปัญญามีค่าเพิ่มขึ้น ค่าตัวแปรประสิทธิภาพของงานจะมีค่าเพิ่มขึ้นตามด้วย ในขณะเดียวกันเมื่อค่าตัวแปรระดับเชาว์ปัญญามีค่าลดลง ค่าตัวแปรประสิทธิภาพของงานจะมีค่าลดลงด้วย

ทั้งค่าความแปรปรวนร่วมและค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน เป็นการหาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร สามารถนำไปใช้เป็นเครื่องมือในการวิเคราะห์เชิงวินิจฉัยได้ อย่างไรก็ตามวิธีการวิเคราะห์ความสัมพันธ์ทั้ง 2 วิธี จะสามารถใช้ได้ก็ต่อเมื่อตัวแปรทั้ง 2 ตัวแปรมีชนิดข้อมูลเป็นข้อมูลเชิงตัวเลขเท่านั้น ยังมีวิธีการทางสถิติอื่นๆ ที่สามารถใช้ในการวัดระดับความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร ให้สามารถเลือกใช้ได้อย่างเหมาะสม เช่น การวิเคราะห์ความแปรปรวนหรืออะโนวา (Analysis of Variance: ANOVA) ซึ่งสามารถใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงตัวเลขและตัวแปรเชิงกลุ่มได้ และการทดสอบไคกำลังสอง (Chi-squared Test) ซึ่งสามารถใช้ทดสอบความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรที่มีชนิดข้อมูลเป็นข้อมูลเชิงกลุ่มได้ เป็นต้น

3.2 การวิเคราะห์กลุ่ม

การวัดระยะห่างระหว่างข้อมูล

การแบ่งกลุ่มข้อมูลแบบเคมีน

การแบ่งกลุ่มแบบลำดับชั้น

การแบ่งกลุ่มเชิงพื้นที่ตามความหนาแน่น

3.3 การวิเคราะห์ความสัมพันธ์

เซตไอเท็ม

กฎความสัมพันธ์