ความแปรปรวน

ความหมายของความแปรปรวน (Variance)

ถ้าเราเก็บข้อมูลจากตัวอย่างจำนวนหนึ่ง แล้วได้คำตอบเหมือนกันหมด คำตอบที่ได้คือค่าคงที่ (constant value) เช่น ถามอายุนักเรียน 5 คน ได้คำตอบว่า 7 ขวบหมดทุกคน ข้อมูลที่ได้คือ 7, 7, 7, 7 และ 7 เป็นข้อมูลคงที่ ไม่มีความแปรปรวนหรือปรวนแปรใด ๆ เลย แต่ถ้าได้คำตอบว่า 1, 4, 7, 10 และ 13 ปี ก็ต้องคิดว่าอายุเฉลี่ยของกลุ่มตัวอย่างนี้เป็นเท่าไร แต่ละคนมีอายุที่แปรเปลี่ยนไปจากอายุเฉลี่ยของกลุ่มมากน้อยเพียงใด เพื่อจะได้ทราบว่าข้อมูลที่ได้รับมีการแปรเปลี่ยนหรือแปรปรวนเท่าใด ดังนั้นความแปรปรวนของข้อมูล จึงเกี่ยวพันกับคะแนนที่เบี่ยงเบนจากคะแนนเฉลี่ย กับจำนวนข้อมูลทั้งหมด

คำว่า "ความแปรปรวน" ทางสถิติ หมายถึงค่าสถิติตัวหนึ่งที่ได้จากผลบวกของกำลังสองของค่าต่าง ๆ ที่เบี่ยงเบนจากค่าเฉลี่ย (mean) หารด้วย N หรือ N-1 เมื่อ N แทนจำนวนของข้อมูลทั้งหมดที่นำมาพิจารณา

ค่าเบี่ยงเบนจากค่าเฉลี่ยบางตัวมีค่าเป็นบวก บางตัวมีค่าเป็นลบ และผลบวกของค่าเบี่ยงเบนเหล่านี้จะเท่ากับศูนย์เสมอ วิธีขจัดปัญหายุ่งยากจากค่าเบี่ยงเบนที่เป็นลบวิธีหนึ่งก็คือการยกกำลังสองค่าเบี่ยงเบนเหล่านี้ทั้งหมด ตัวอย่างเช่น ข้อมูลชุดหนึ่งคือ 1, 4, 7, 10 และ 13 หาค่าเฉลี่ยได้เท่ากับ 7 เมื่อนำคะแนนแต่ละตัวไปลบออกจาก 7 จะได้ค่าเบี่ยงเบนเท่ากับ − 6, − 3, 0, + 3, และ + 6 ตามลำดับ ยกกำลังสองค่าเบี่ยงเบนเหล่านี้จะได้ 36, 9, 0, 9, และ 36 นำค่าที่ยกกำลังสองแล้วมาบวกกันได้เท่ากับ 90 ถ้าเอา 90 หารด้วย 5 จะได้ 18 แต่ถ้าเอา 90 หารด้วย 5 − 1 หรือ 4 จะได้ 22.50 ค่า 18 หรือ 22.50 นี้แหละที่เรียกว่าค่าความแปรปรวนของข้อมูลชุดนี้

การหาค่าความแปรปรวน

ด้วยวิธีการที่กล่าวมาแล้ว เมื่อนำมาเขียนเป็นสูตร โดยกำหนดให้ s^2 แทนความแปรปรวน จะได้

แบบที่หนึ่ง : $s^2 = \frac{\Sigma(X-\bar{X})^2}{N}$

แบบที่สอง : $s^2 = \frac{\Sigma(X-\bar{X})^2}{N-1}$

เมื่อพิจารณาความแตกต่างที่เป็นสาระสำคัญ ระหว่างความแปรปรวนแบบที่หนึ่งกับแบบที่สอง จะเกี่ยวโยงไปถึงค่าจากกลุ่มตัวอย่างเรียกว่าค่าประมาณ (estimates) หรือค่าสถิติ (statistic) และค่าจากประชากรเรียกว่า **พารามิเตอร์** (parameters) ค่าความแปรปรวนทั้งแบบที่หนึ่งและแบบที่สอง ต่างก็เป็นค่าประมาณของค่าความแปรปรวนจากประชากร (population variance) เมื่อหาร $\Sigma(X-\bar{X})^2$ ด้วย N ได้ค่าที่เรียกว่าค่าประมาณที่เอนเอียงของ σ^2 (a biased estimate of σ^2) ค่าประมาณตัวนี้มีแนวโน้มที่จะมีค่าน้อยกว่า σ^2 แต่เมื่อหาร $\Sigma(X-\bar{X})^2$ ด้วย N-1 จะได้ค่าที่เรียกว่าค่าประมาณที่ไม่เอนเอียงของ (an unbiased estimate of σ) ค่าประมาณทั้งสองค่าที่ได้แทนด้วยสัญลักษณ์ s^2 เพราะไม่ใช่ค่าพารามิเตอร์ เนื่องจากหาค่ามาจากคะแนนที่เบี่ยงเบนจาก $\bar{X}$ มิได้หามาจากคะแนนที่เบี่ยงเบนจากคะแนนเฉลี่ยของประชากร หรือ μ

ดังนั้นในกรณีที่ต้องการหาค่า σ^2 โดยตรง ควรใช้สูตร

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N_p}$$

เมื่อ μ คือค่าเฉลี่ยของประชากร (population mean) และ N_p คือจำนวนสมาชิกที่เป็นประชากรทั้งหมด

ความสำคัญของ N - 1 ที่มีต่อการคำนวณค่าความแปรปรวน

เมื่อ N คือ จำนวนสมาชิกส่วนหนึ่งของประชากรที่นำมาพิจารณาด้วยวิธีการต่าง ๆ N - 1 ก็คือจำนวนของค่าที่เบี่ยงเบนจากค่าเฉลี่ยได้อย่างอิสระเสรี สมมติว่า มีข้อมูล 3 จำนวน คือ 7, 8, และ 15 ค่าเฉลี่ยคือ 10 ค่าเบี่ยงเบนจากค่าเฉลี่ยก็คือ – 3, – 2, และ + 5 จะเห็นว่าผลบวกของค่าเฉลี่ยเท่ากับ $(-3) + (-2) + (5) = 0$ พิจารณาค่าเบี่ยงเบน 3 จำนวนนี้จะพบว่า ถ้ากำหนดค่าเบี่ยงเบนสองจำนวนใด ๆ แล้ว ค่าเบี่ยงเบนจำนวนที่สามจะต้องคงที่จะแปรเปลี่ยนอย่างอิสระไม่ได้ เพราะบังคับให้ผลรวมต้องเท่ากับศูนย์ ดังนั้นจากข้อมูล 3 จำนวน จะแปรเปลี่ยนอย่างอิสระได้ 3 – 1 หรือ 2 จำนวน ถ้ามีข้อมูล 5 จำนวน ก็จะแปรเปลี่ยนอย่างอิสระได้ 5 – 1 หรือ 4 จำนวน จำนวนค่าที่สามารถแปรเปลี่ยนได้อย่างอิสระเช่นนี้เรียกว่า "ชั้นแห่งความเป็นอิสระ" (degrees of freedom) จึงกล่าวได้ว่าค่า $\Sigma(X-\bar{X})^2$ มีความสัมพันธ์กับค่าชั้นแห่งความเป็นอิสระ N - 1 เพราะว่าเมื่อยกกำลังสองค่าเบี่ยงเบนแต่ละจำนวนจะมี N - 1 จำนวน จาก N จำนวน ที่สามารถแปรเปลี่ยนได้อย่างอิสระ จากข้อมูลข้างต้น เมื่อยกกำลังสองค่าเบี่ยงเบนทั้ง 3 จำนวนแล้วบวกกันจะได้ 9 + 4 + 25 = 38 เราจะกำหนดค่าตัวที่ 1 กับตัวที่ 2 เป็นเท่าใดก็ได้ แต่ตัวที่ 3 ต้องเป็นค่าแน่นอน ที่จะทำให้ผลรวมเท่ากับ 38 ค่าตัวที่ 3 จึงแปรเปลี่ยนอย่างอิสระไม่ได้

ความแปรปรวนและความเบี่ยงเบนมาตรฐาน

ค่าความแปรปรวนเป็นค่าสถิติที่มีหน่วยเป็นกำลังสอง (squared units) ถ้า $X - \bar{X}$ เป็นความเบี่ยงเบนที่มีหน่วยเป็น นิ้ว แล้วค่า $(X - \bar{X})^2$ จะเป็นความเบี่ยงเบนที่มีหน่วยเป็นตารางนิ้ว การวัดหลายแบบที่ไม่ประสงค์จะให้มีหน่วยเป็นกำลังสอง แต่ต้องการหน่วยปกติที่ใช้วัดทั่ว ๆ ไป จึงต้องเปลี่ยนหน่วยความแปรปรวนเสียใหม่ ด้วยการถอดรากที่สอง ค่าที่ได้รับเป็นค่าสถิติที่เรียกว่า "ความเบี่ยงเบนมาตรฐาน" (standard deviation) เขียนเป็นสูตรแบบต่าง ๆ ได้ดังนี้

แบบที่หนึ่ง: $s = \sqrt{\dfrac{\Sigma(X-\bar{X})^2}{N}}$

แบบที่สอง: $s = \sqrt{\dfrac{\Sigma(X-\bar{X})^2}{N-1}}$

แบบที่สาม: $s = \sqrt{\dfrac{\Sigma X^2 - N\bar{X}^2}{N-1}}$

วิธีหาความแปรปรวนจากสูตรแบบที่สาม เริ่มจากหาผลบวกของกำลังสองของคะแนนดิบแต่ละตัวแล้วลบด้วยค่าเฉลี่ยของกำลังสองที่คูณกับ N แล้วหารด้วย N-1 เช่น มีคะแนนดิบ 5 จำนวน คือ 1, 4, 7 10 และ 13 ค่าเฉลี่ยคือ 7 กำลังสองของคะแนนแต่ละตัวคือ 1, 16, 49, 100 และ 169 ผลบวกของค่ากำลังสองเหล่านี้คือ 335 จะได้ค่าความแปรปรวน ดังนี้

$$s^2 = \frac{\Sigma X^2 - N\bar{X}^2}{N-1}$$

$$= \frac{335 - 5(7)^2}{5-1} = 22.50$$

ถ้าไม่ต้องการเขียนสูตรในรูปของ $\bar{X}$ ก็สามารถเปลี่ยน $\bar{X}$ เป็น $\Sigma X/N$ แล้วนำไปแทนค่าในสูตรข้างบนนี้ จะได้ค่าความแปรปรวน คือ

$$s^2 = \frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}$$

การบวกหรือการคูณความแปรปรวนด้วยค่าคงที่

ถ้าเอาค่าคงที่ใด ๆ บวกคะแนนทุกตัวในข้อมูลชุดใดชุดหนึ่ง ค่าความแปรปรวนของข้อมูลชุดนั้นไม่เปลี่ยนแปลง จากตัวอย่างที่กล่าวมาแล้ว ถ้าเอา 5 บวกคะแนนดิบ 1, 4, 7, 10 และ 13 จะได้ 6, 9, 12, 15, และ 18 ค่าเฉลี่ยของข้อมูลเดิมคือ 7 จะเพิ่มเป็น 7 + 5 หรือ 12 ค่าเบี่ยงเบนจากค่าเฉลี่ย 12 ของคะแนนแต่ละตัวก็คือ – 6, – 3, 0, + 3 และ + 6 เช่นเดิม ความแปรปรวนของข้อมูลชุดใหม่ที่บวกค่าคงที่ไปแล้ว ยังคงเป็น 22.50 เท่าเดิม

ถ้าเอาค่าคงที่ใด ๆ คูณกับคะแนนทุกตัว ในข้อมูลชุดใดชุดหนึ่ง ค่าความแปรปรวนของข้อมูลใหม่นี้จะเท่ากับกำลังสองของค่าคงที่คูณกับค่าความแปรปรวนเดิม กล่าวคือ

ถ้าค่าคงที่เหลือ c ค่าความแปรปรวนจะเท่ากับ c^2s^2 เช่น ถ้าเอา 3 คูณกับคะแนน 1, 4, 7, 10 และ 13 ทุกตัวแล้วจะได้ความแปรปรวนเท่ากับ 9 (22.50) = 202.50

ความแปรปรวนรวม (pooled variance)

ในกรณีที่มีกลุ่มตัวอย่าง 2 กลุ่ม ค่าสถิติที่ได้คือ ค่าเฉลี่ยของกลุ่มหนึ่ง ($\bar{X}_1$) ค่าเฉลี่ยของกลุ่มสอง ($\bar{X}_2$) จำนวนสมาชิกกลุ่มหนึ่ง (N_1) จำนวนสมาชิกกลุ่มสอง (N_2) ค่าความแปรปรวนของกลุ่มหนึ่ง(s_1^2)และค่าความแปรปรวนของกลุ่มสอง(s_2^2) จึงเกิดความจำเป็นจะต้องหาค่าความแปรปรวนของกลุ่มตัวอย่างทั้งสองรวมกัน เพื่อให้ได้ค่าประมาณที่ไม่เอนเอียงของความแปรปรวนของประชากร (an unbiased estimate of the population variance) ค่าประมาณนี้เรียกว่าค่าความแปรปรวนรวม ได้จากการเอาผลบวกของกำลังสองของค่าเบี่ยงเบนจากค่าเฉลี่ยของกลุ่มตัวอย่างทั้งสองมาบวกกันหรือมารวมกัน แล้วหารด้วยชั้นแห่งความเป็นอิสระของทั้งหมด เขียนเป็นสูตรเพื่อการคิดคำนวณได้ดังนี้

$$s_t^2 = \frac{\overset{N1}{\Sigma}\Sigma(X-\bar{X}_1)^2 + \overset{N2}{\Sigma}\Sigma(X-\bar{X}_2)^2}{N_1 + N_2 - 2}$$

$$s_t^2 = \frac{s_1^2(N_1-1)+s_2^2(N_2-1)}{N_1 + N_2 - 2}$$

ชั้นแห่งความเป็นอิสระของกลุ่มหนึ่งคือ $N_1 - 1$ และชั้นแห่งความเป็นอิสระของกลุ่มสองคือ $N_2 - 1$
ดังนั้นชั้นแห่งความเป็นอิสระของกลุ่มตัวอย่างทั้งสองรวมกันจึงมีค่าเท่ากับ $N_1 + N_2 - 2$
ค่าความแปรปรวนรวมที่ได้นี้ อาจแทนด้วยสัญลักษณ์ S_p^2 หรือ S_t^2 ก็ได้

การแปลความหมายและการนำไปใช้

สมมติว่ากลุ่มตัวอย่างแรก มีค่าความแปรปรวนเท่ากับ 4 กลุ่มตัวอย่างที่สองมีค่าความแปรปรวนเท่ากับ 9 แต่ละกลุ่มมีจำนวนตัวอย่างกลุ่มละ 11 คน เราต้องการวิเคราะห์ข้อมูลของทั้งสองกลุ่มรวมกันว่า มีลักษณะแปรเปลี่ยนอย่างไร จะใช้ความแปรปรวนของกลุ่มใดกลุ่มหนึ่ง มาคิดคำนวณไม่ได้ จึงจำเป็นต้องหา "ความแปรปรวนรวม" ด้วยการแทนค่าในสูตรข้างต้น ได้ความแปรปรวนรวมเท่ากับ 6.5 จะเห็นว่าความแปรปรวนรวมมีค่าอยู่ระหว่างความแปรปรวนของทั้งสองกลุ่ม การใช้สูตรดังกล่าวเมื่อทราบค่าความแปรปรวนและจำนวนของกลุ่มตัวอย่างในแต่ละกลุ่ม จะช่วยให้หาค่าความแปรปรวนได้รวดเร็วและสะดวกยิ่งขึ้น

ความแปรปรวนร่วม (covariance)

บางครั้งการพิจารณาความสัมพันธ์ของข้อมูลจากกลุ่มตัวอย่างสองกลุ่ม จะมิใช่ลักษณะของการรวมความแปรปรวนเข้าด้วยกัน แต่จะพิจารณาในลักษณะของตัวแปรร่วม (co-variate of concomitant variable) เช่นตัวแปร X และตัวแปร Y มีความสัมพันธ์ร่วมกันอย่างไรบ้าง จึงเกี่ยวโยงไปถึงความแปรปรวนของ X กับ ความแปรปรวนของ Y ว่ามีความสัมพันธ์ร่วมกันอย่างไร อันเป็นเหตุให้ต้องมีการวิเคราะห์ความแปรปรวนร่วมกันต่อไป

การหาค่าความแปรปรวนร่วม ได้จากการเอาผลคูณที่เกิดจากคะแนนเบี่ยงเบนจากค่าเฉลี่ยของกลุ่ม X และกลุ่ม Y มาบวกกันแล้วหารด้วย N - 1 เขียนเป็นสูตรจะได้

$$s_{xy}^2 = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{N-1}$$

ถ้าเป็นข้อมูลของประชากร สัญลักษณ์ที่ใช้แทนความแปรปรวนร่วมของประชากร (population covariance) คือ σ_{xy}^2 หาค่าได้ในทำนองเดียวกันกับที่กล่าวมาแล้วคือ

$$\sigma_{xy}^2 = \frac{\Sigma(X-\mu_x)(Y-\mu_y)}{N}$$

สัญลักษณ์ที่ใช้แทนความแปรปรวนร่วม นอกจากที่กล่าวมาแล้ว จะใช้ในรูปตัวย่อว่า Cov หรือ Cov_{xy} ก็ได้ ถ้าต้องการหาสัมประสิทธิ์สหสัมพันธ์ที่มีความแปรปรวนร่วมกันระหว่างข้อมูล X และข้อมูล Y สามารถนำค่าความแปรปรวนร่วม และความแปรปรวนของข้อมูล X กับความแปร

ปรวนของข้อมูล Y มาแทนค่าในสูตรต่อไปนี้ได้

$$r_{xy} = \frac{s_{xy}^2}{\sqrt{s_x^2 s_y^2}}$$

รายละเอียดของการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบต่าง ๆ จะกล่าวถึงต่อไปในเรื่องสหสัมพันธ์ (correlation)

การแปลความหมายและการนำไปใช้

สมมติว่านักเรียนกลุ่มหนึ่งทำแบบทดสอบสองชุด คือ การคิดอย่างมีเหตุผล (แทนด้วย X) และการใช้ภาษา (แทนด้วย Y) โดยที่ความแปรปรวนของคะแนน X เท่ากับ 9 ความแปรปรวนของคะแนน Y เท่ากับ 16 ตัวแปรทั้งสองนี้มีค่าสัมประสิทธิ์สหสัมพันธ์ (r_{xy}) เท่ากับ 0.5 เมื่อแทนค่าดังกล่าวในสูตรข้างต้น จะได้ความแปรปรวนร่วมเท่ากับ 6 แสดงให้เห็นว่าคะแนนการคิดอย่างมีเหตุผล และการใช้ภาษามีการกระจายหรือการเปรเปลี่ยนร่วมกันเป็น 6 ถ้าจะให้ได้คำตอบแน่ชัดว่า ความแปรปรวนร่วมของข้อมูลชุดนี้มีค่ามากน้อยเพียงใด จะสรุปผลไปยังข้อมูลดิบที่ได้มาในลักษณะใด ใช้วิธีการทางสถิติที่เรียกว่า "การวิเคราะห์ความแปรปรวนร่วม" ซึ่งจะได้กล่าวถึงต่อไป

บังอร ภูวภิรมย์ขวัญ

บรรณานุกรม

บังอร ภูวภิรมย์ขวัญ **สถิติประยุกต์ทางการศึกษา** หวีกิจการพิมพ์ 2523 : 215

Ferguson, George A. **Statistical Analysis in Psychology and Education.** 2 nd. McGraw-Hill, Inc., 1981 : 549.