วิธีทดสอบความเป็นเนื้อเดียวกัน และความเสถียรของตัวอย่าง

ศิริวรรณ ศิลปสกุลสุข
กานดา โกมลวัฒนวัฒ

การทดสอบความชำนาญของห้องปฏิบัติการ (Laboratory Proficiency Testing) ตาม ISO/IEC Guide 43-1 มีอยู่หลายรูปแบบ และเทคนิคการเปรียบเทียบระหว่างห้องปฏิบัติการ (Interlaboratory Testing Schemes) เป็นรูปแบบหนึ่งเพื่อวัตถุประสงค์ในการตรวจสอบความสามารถของห้องปฏิบัติการ

การเปรียบเทียบผลระหว่างห้องปฏิบัติการจะใช้ตัวอย่างเดียวกันซึ่งส่งไปยังห้องปฏิบัติการต่างๆ ที่เข้าร่วมกิจกรรมการทดสอบความชำนาญเพื่อดำเนินการภายใต้เงื่อนไขที่ได้ตกลงกันไว้ก่อน และผู้ส่งตัวอย่างจะประเมินผลการวิเคราะห์ทดสอบของห้องปฏิบัติการเหล่านั้น ดังนั้นต้องแน่ใจว่าตัวอย่างที่ใช้เปรียบเทียบผลระหว่างห้องปฏิบัติการนั้นมีความเป็นเนื้อเดียวอย่างสม่ำเสมอ (homogeneity) เพื่อที่จะแน่ใจได้ว่าผลของห้องปฏิบัติการที่เบี่ยงเบนไปนั้นไม่ได้เกิดจากความต่างของตัวอย่าง ในที่นี้จะขอกล่าวถึงวิธีการทดสอบความเป็นเนื้อเดียวกันของตัวอย่าง (test sample) จากแนวทางของ The Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories เป็นการยกตัวอย่างที่ใช้ขบวนการทางสถิติซึ่งประกอบด้วยข้อแนะแนวจาก ISO Guide 35 และเป็นตัวอย่างข้อมูล ที่ได้จากโปรแกรมการเปรียบเทียบความชำนาญระหว่างห้องปฏิบัติการที่ทดสอบทางด้านสิ่งแวดล้อม โดยมีขั้นตอนดังนี้

1. สุ่มเลือกตัวอย่างที่จะทดสอบ 10 ตัวอย่างจากตัวอย่างทั้งหมดซึ่งพร้อมจะแจกไปยังผู้ทำการทดสอบ

2. แยกเตรียมตัวอย่างที่จะทำการทดสอบแต่ละตัวอย่างโดยแบ่งออกเป็นสองส่วน จะได้ 20 ส่วน

3. ทำการวิเคราะห์ทั้ง 20 ส่วนโดยให้ผลการทดสอบทั้งสองค่าในแต่ละตัวอย่างมีค่าต่างกันไม่เกินค่า repeatability ของแต่ละการทดสอบนั้นๆ (repeatability condition)

4. ประมาณค่า ความแปรปรวนระหว่างตัวอย่าง (Between-group Variability) S_s^2 และค่าแปรปรวนจากการวิเคราะห์ (analytical variance) หรือความแปรปรวนภายในกลุ่ม (Within-group Variability) S_a^2 โดยใช้ การวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA แบบทางเดียว) และไม่ต้องตัด outliers

5. รายงานค่า $\overline{X}$, S_s, S_a, n และผลของ F-test ถ้าค่าที่คำนวณได้ของ F น้อยกว่า ค่าวิกฤติของ F (ซึ่งได้จากการเปิดตาราง) แสดงว่าตัวอย่างที่ใช้ทดสอบมีความเป็นเนื้อเดียวกัน

6. ถ้าค่าที่คำนวณได้ของ F มากกว่าค่าวิกฤติของ F (จากตาราง) หรือที่ระดับความเชื่อมั่น 95% มีค่านัยสำคัญทางสถิติ (significant) น้อยกว่า 0.05 แสดงว่าตัวอย่างที่ใช้ทดสอบไม่เป็นเนื้อเดียวกันในทางสถิติ ให้เปรียบเทียบค่าประมาณของ S_s กับค่าช่วงประเมิน (evaluation interval) $3\sigma_t$ โดย σ_t เป็นค่าความเบี่ยงเบนระหว่างตัวอย่าง ซึ่งได้จากการทำการเปรียบเทียบผลระหว่างห้องปฏิบัติการในอดีต ถ้าค่า S_a น้อยกว่า 0.1 ($3\sigma_t$) หรือ 10% ของค่าช่วงประเมิน ตัวอย่างอาจจะมีความเป็นเนื้อเดียวกันเพียงพอที่จะใช้ในการทดสอบ

7. ถ้าตัวอย่างมีความเป็นเนื้อเดียวกันจากการทดสอบ F-test ค่าประมาณที่ดีที่สุดของ σ_i (ค่าเบี่ยงเบนมาตรฐานของประชากรจริง) ก็คือค่าความเบี่ยงเบนมาตราฐาน 20 การทดสอบ และถ้าตัวอย่างไม่เป็นเนื้อเดียวกันจากการทดสอบ F-test แต่ค่า S_s น้อยกว่า 0.1 ($3\sigma_t$) หรือ 10% ของค่าช่วงประเมิน ดังนั้นตัวอย่างที่ยอมรับให้ใช้ทำการทดสอบได้จำเป็นจะต้องประมาณค่า σ_i ค่าประมาณที่ดีที่สุดของ σ_i คือค่า variance กลาง (pooled variance)

$$\sigma_i^2 = \sigma_s^2 + \sigma_a^2$$

โดยประมาณจาก $S_i^2 = S_s^2 + S_a^2$

ค่าประมาณนี้จะใช้ในการวินิจฉัยครั้งต่อไป ซึ่งจะรวมระดับความเชื่อมั่น หรือการประมาณค่าความไม่แน่นอน ค่านี้จะใช้ในการทดสอบความมีนัยสำคัญของการเปลี่ยนแปลงใดๆ ที่เกิดจากความไม่เสถียร (instability) ได้อีกด้วย

ตัวอย่าง การทดสอบ Homogeneity ของสารละลายเกลือคลอไรด์ที่เตรียมขึ้นในห้องปฏิบัติการ สุ่มทำการทดสอบ 10 ตัวอย่างให้ผลดังตารางต่อไปนี้

Sample	Dilution # 1	Dilution # 2
1	180.41	180.69
2	180.30	180.58
3	180.26	180.83
4	180.44	180.72
5	180.76	180.62
6	181.24	180.98
7	181.03	181.32
8	180.57	180.12
9	180.15	180.39
10	181.26	181.00

(หมายเหตุ : ปริมาณคลอไรด์ในน้ำ ; หน่วย mg/L)

ขั้นตอนในการวิเคราะห์ผลทางสถิติ

1. หาค่าเฉลี่ยและค่าความเบี่ยงเบนของ 20 การทดสอบ $\overline{X}_0$ = 180.6835, S_0 = 0.3671

2. วิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA แบบทางเดียว)

ANOVA

Source of Variation	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.0309	9	0.2257	4.261	0.0168
Within Groups	0.5296	10	0.0530		
Total	2.5605	19			

3. ค่า S_s^2 = 0.2257 และ S_a^2 = 0.0530

4. ค่าเบี่ยงเบนมาตรฐานระหว่างตัวอย่าง S_s คือ $\sqrt{MSTrt} = \sqrt{0.2256}$ = 0.4750 และค่าเบี่ยงเบนมาตรฐานจากการวิเคราะห์ S_a คือ $\sqrt{MSE} = \sqrt{0.0530}$ = 0.2302

5. ค่า F_{cal} = 4.261 มากกว่าค่าวิกฤติของ F คือ $F_{0.05,9,10}$ = 3.02 หรือพิจารณาจากค่านัยสำคัญทางสถิติ (Significant) เท่ากับ 0.0168 ซึ่งน้อยกว่า 0.05 แสดงว่าตัวอย่างที่ใช้ทดสอบไม่เป็นเนื้อเดียวกันในทางสถิติ

6. เมื่อตัวอย่างไม่เป็นเนื้อเดียวกันอย่างมีนัยสำคัญในเชิงสถิติ จำเป็นต้องเปรียบเทียบขนาดของความไม่เป็นเนื้อเดียวกันกับขนาดของช่วงประเมินที่ความเข้มข้นเดียวกัน ค่าประมาณของค่าเบี่ยงเบนระหว่าง ตัวอย่าง S_s คือ 0.475 สมมติว่าการทดสอบนี้ถูกประเมินด้วยสถิติ Z และค่าความเบี่ยงเบนมาตรฐานระหว่างห้องปฏิบัติการที่ความเข้มข้นเดียวกันนี้ σ_t มีค่าประมาณ 1.8 mg/L ดังนั้น $3\sigma_t$ = 5.4 จะได้อัตราส่วนของความเบี่ยงเบนระหว่างตัวอย่างต่อค่าช่วงประเมินเป็น 0.475/5.4 = 0.088 = 8.8% น้อยกว่า 10% ซึ่งค่านี้นับว่ามีความสัมพันธ์น้อยเมื่อเทียบกับการใช้เป็นตัวอย่างทดสอบ ดังนั้นยังถือว่าตัวอย่างมีความเหมาะสมในการใช้ทดสอบ เมื่อตัวอย่างนี้ไม่เป็นเนื้อเดียวกัน การประมาณค่าที่ดีที่สุดของ σ คือการใช้ค่า variance รวม (pooled variance) ; $\sigma_s^2 + \sigma_a^2$, โดยได้จากค่าประมาณของตัวอย่าง

$$S = \sqrt{(S_s^2 + S_a^2)} = \sqrt{(0.2256 + 0.530)}$$

$$= 0.5278$$

นอกจากนี้ตัวอย่างที่ใช้เปรียบเทียบผลระหว่างห้องปฏิบัติการนั้นผู้เตรียมตัวอย่างต้องหาวิธีตรวจสอบความเสถียร (stability) ด้วยเพื่อเป็นการประกันว่าตัวอย่างจะไม่มีการเปลี่ยนแปลงในช่วงเวลาที่ใช้ทดสอบความชำนาญของห้องปฏิบัติการ

การหาความเสถียรของตัวอย่างค่อนข้างง่ายกว่าการหาความเป็นเนื้อเดียวกัน การหาความเสถียรนั้นจะพิจารณาความแตกต่างค่าเฉลี่ย มากกว่าที่จะเป็นการประมาณค่าความเบี่ยงเบนสัมพันธ์ ดังนั้นการใช้การวิเคราะห์ทางสถิติและการทดสอบแบบไม่ซับซ้อนจึงเพียงพอสำหรับการทดสอบความเสถียรและสามารถใช้ผลจากการทดสอบความเป็นเนื้อเดียวกันในตอนต้นกับผลของตัวอย่างจากการทดสอบความเสถียรโดยทำการทดสอบความแตกต่างของค่าเฉลี่ยของสองกลุ่มด้วย t-test อย่างไรก็ตามไม่ควรใช้การทดสอบทางสถิตินี้เป็นเกณฑ์เดียวสำหรับการทดสอบความเสถียร เนื่องจากการเบี่ยงเบนจากค่าเฉลี่ยอย่างมีนัยสำคัญทางสถิติอาจจะไม่ได้หมายความว่าตัวอย่างนั้นไม่เสถียรเพียงพอที่จะใช้ในการทดสอบความชำนาญ ในกรณีที่วิธีทดสอบที่ใช้ความถูกต้องสูงมาก อาจจะทำให้ช่วงความเชื่อมั่นแคบและมีข้อจำกัดในการใช้ตัวอย่างมากเกินความเป็นจริง เมื่อมีผลกระทบจากการประเมินเข้ามาเกี่ยวข้องก็จำเป็นที่จะต้องใช้ช่วงประเมินเข้ามาเป็นตัวตัดสินความสำคัญของการเบี่ยงเบนไปจากค่าเฉลี่ยที่เกิดขึ้น ความแตกต่างระหว่างค่าเฉลี่ยตัวอย่างที่ทดสอบความเสถียรกับค่าเฉลี่ยของผลจากการทดสอบความเป็นเนื้อเดียวกัน ควรนำมาเปรียบเทียบกับความกว้างของช่วงประเมิน ถ้าต่างกันน้อยกว่า 10% ของช่วงประเมินอาจกล่าวได้ว่าตัวอย่างมีความเสถียรเพียงพอที่จะใช้เพื่อทดสอบความชำนาญ

ขั้นตอนในการทดสอบความเสถียร ในที่นี้จะขอกล่าวถึงวิธีการทดสอบของตัวอย่างที่ได้จากรายการทดสอบความชำนาญสำหรับห้องปฏิบัติการที่ทดสอบทางสิ่งแวดล้อมมีขั้นตอนการดำเนินการดังต่อไปนี้

1. สุ่มเลือก 5 ตัวอย่างทดสอบจากตัวอย่างทั้งหมดที่เตรียมแจกให้ผู้ร่วมรายการทดสอบความชำนาญ

2. แยกเตรียมแต่ละตัวอย่างทดสอบ และแบ่งออกมาเพียง 1 ส่วนจากแต่ละตัวอย่าง

3. วิเคราะห์ทั้ง 5 ส่วนทดสอบโดยใช้เงื่อนไขเดียวกับการทำซ้ำ (repeatability)

4. คำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของการทดสอบทั้ง 5 ส่วน โดยไม่ต้องตัด outliers

5. ทำการทดสอบความแตกต่างของค่าเฉลี่ยของสองกลุ่ม (กลุ่มตัวอย่างการทดสอบความเป็นเนื้อเดียวกันกับกลุ่มตัวอย่างการทดสอบความเสถียร) ด้วย t-test แต่ถ้าค่าที่คำนวณได้ของ t น้อยกว่าค่าวิกฤตของ $|t|$ แสดงว่าตัวอย่างที่ใช้ทดสอบยังไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญ

6. แต่ถ้าค่าที่คำนวณได้ของ t มากกว่าค่าวิกฤตของ $|t|$ แสดงว่าตัวอย่างที่ใช้ทดสอบไม่เสถียรในทางสถิติ ให้เปรียบเทียบค่าความแตกต่างของค่าเฉลี่ยระหว่างตัวอย่างที่ทดสอบเสถียรกับตัวอย่างที่ทดสอบความเป็นเนื้อเดียวกัน กับค่าช่วงประเมิน $3\sigma_t$ ถ้าค่าความแตกต่างของค่าเฉลี่ยระหว่างตัวอย่างที่ทดสอบเสถียรกับตัวอย่างที่ทดสอบความเป็นเนื้อเดียวกันน้อยกว่า 0.1 ($3\sigma_t$) หรือ 10% ของค่าช่วงประเมิน ความไม่เสถียรนี้ไม่กระทบต่อการประเมิน

ตัวอย่าง การทดสอบ Stability ของสารละลายเกลือคลอไรด์ที่เตรียมขึ้นในห้องปฏิบัติการ สุ่มทำการทดสอบให้ผลดังตารางต่อไปนี้

Sample	Result
1	180.22
2	180.17
3	180.35
4	180.42
5	180.15

(หมายเหตุ : ปริมาณคลอไรด์ในน้ำ ; หน่วย mg/L)

ขั้นตอนในการวิเคราะห์ผลทางสถิติ

1. ได้ค่าเฉลี่ย $\overline{X}_1$ = 180.262 S_1 = 0.1178 และจากการทดสอบความเป็นเนื้อเดียวกัน (จากตัวอย่างจำนวน 20 ตัวอย่าง)

ได้ค่าเฉลี่ย $\overline{X}_0$ = 180.6835

ค่า เบี่ยงเบนมาตรฐาน S_0 = 0.5278

2. ทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ยของสองประชากร t-test ได้ค่า t_{cal} = 3.2342 ที่ระดับความเป็นอิสระ (degrees of freedom) df = 23 และที่ระดับความเชื่อมั่น 95% ค่าวิกฤตของ t คือ $t_{0.025,23}$ = 2.07 นั่นแสดงว่าค่า t_{cal} มากกว่าค่าวิกฤตของ $t_{0.025,23}$ แสดงว่าตัวอย่างที่ใช้ทดสอบไม่เสถียรในทางสถิติ

3. ความแตกต่างของค่าเฉลี่ยระหว่างตัวอย่างที่ทดสอบเสถียรกับตัวอย่างที่ทดสอบความเป็นเนื้อเดียวกัน (180.6835 - 180.262 = 0.4215) น้อยกว่า 0.1 ($3\sigma_t$) = 0.54 นั่นแสดงว่า ความไม่เสถียรนี้ไม่กระทบต่อการประเมิน

วิธีตรวจสอบความเป็นเนื้อเดียวกันและความเสถียรที่กล่าวไปแล้วนั้นเป็นข้อเสนอแนะการออกแบบการทดสอบความเป็นเนื้อเดียวกันและความเสถียร ผู้เตรียมตัวอย่างอาจจะประยุกต์วิธีการตรวจสอบที่ใช้อยู่เพื่อเป็นการประกันความเป็นเนื้อเดียวกันและความเสถียรของตัวอย่างที่แจกออกไป โดยอาศัยประสบการณ์ที่เคยสัมผัสกับตัวอย่างที่ใช้เฉพาะในการทดสอบความชำนาญและข้อมูลที่มีอยู่อย่างเพียงพอ

เอกสารอ้างอิง

Miller, JC. and Miller, JN. **Statistics and chemometrics for analytical chemistry.** 4th ed. London : Pearson Education. 2000. p 44-48, 58-64.

The International Organization for Standardization/The International Electrotechnical Commission. **Proficiency testing by interlaboratory comparisons-Part 1** : development and operation of proficiency testing schemes. ISO/IEC Guide 43-1.1997.

Thomson, M. and Wood, R. The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories. **Pure and Applied Chemistry**. 1993. vol. 65. no.9.p. 2123-2144.

กัลยา วานิชย์บัญชา. **การวิเคราะห์สถิติ : สถิติเพื่อการตัดสินใจ.** พิมพ์ครั้งที่ 5. กรุงเทพมหานคร : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย. 2544. หน้า 140-145.