สถิติกับเคมีวิเคราะห์

ภัทรา ปัญญวัฒนกิจ

สถิติเป็นวิทยาศาสตร์แขนงหนึ่งที่เกี่ยวข้องกับหลักการและระเบียบวิธีการทางสถิติซึ่งประกอบด้วยกระบวนการที่สำคัญคือ การเก็บรวบรวม การนำเสนอ การวิเคราะห์ข้อมูล การตีความหมายและการสรุปข้อมูล เพื่อประโยชน์ในการตัดสินใจที่ถูกต้อง ในแง่ของตัวเลข สถิติ หมายถึง ตัวเลขที่เกิดจากการวัด การคิดคำนวณ จากข้อมูล ที่รวบรวมมาตามหลักเกณฑ์ เช่น คะแนนเฉลี่ยของวิชาคณิตศาสตร์ในการสอบเข้ามหาวิทยาลัยประจำปีการศึกษา 2538 รายได้โดยเฉลี่ยต่อปีของคนไทยในปี 2541 ยอดรวมของสินค้าทางการเกษตรที่ส่งออกในปี 2537 ซึ่งตัวเลขเหล่านี้ล้วนเป็นตัวเลขทางสถิติทั้งสิ้น ปัจจุบันวิชาสถิติเข้ามามีบทบาทในชีวิตประจำวันเป็นอย่างยิ่งทั้งทางด้านธุรกิจ วิทยาศาสตร์ การแพทย์ และสังคมศาสตร์เป็นต้น ไม่ว่าจะเป็นการเผยแพร่ในหนังสือพิมพ์ วิทยุ โทรทัศน์ หรือ วารสารต่าง ๆ ซึ่งจะมีการยกตัวอย่างสถิติมาอ้างอิงอยู่เสมอ สถิติจึงได้รับการยอมรับให้เป็นภาษาสากลของการเผยแพร่ข้อมูลข่าวสาร และผลงานการวิจัยต่าง ๆ ดังที่ H.G. Well นักปราชญ์ผู้มีชื่อเสียงในสมัยศตวรรษที่ 19 ได้กล่าวไว้ว่า "ในอนาคตข้างหน้าสถิติจะเป็นองค์ประกอบที่สำคัญประการหนึ่งของชนชาติต่าง ๆ เช่นเดียวกับการอ่านและการเขียน"

เคมีวิเคราะห์เป็นวิทยาศาสตร์สาขาหนึ่งที่มีความสำคัญต่อการดำรงชีวิตเป็นอย่างมาก เป็นต้นว่า การวิเคราะห์ โลหะหนักต่าง ๆ เช่น สารหนู ตะกั่ว ปรอท ในอาหารกระป๋อง ปริมาณโบรอนในตัวอย่างดิน สารบอแรกซ์ในเนื้อหมู เป็นต้น ซึ่งห้องปฏิบัติการมักพบปัญหาอยู่เสมอ เช่น การหาปริมาณโบรอนในตัวอย่างดินจำนวน 4 ครั้ง ได้ผลการวิเคราะห์ดังนี้ 24.69, 24.73, 24.72 และ 25.39 มิลลิกรัมต่อกิโลกรัม การวิเคราะห์หาปริมาณตะกั่วในอาหารกระป๋องด้วยเครื่องอะตอมมิกแอบซอร์พชัน สเปคโทร โฟโตมิเตอร์ กับเครื่อง ยูวี- วิซิเบิล สเปคโทรโฟโตมิเตอร์ ได้ปริมาณตะกั่ว 1.23 และ1.54 มิลลิกรัม ต่อกิโลกรัม เป็นต้น พบว่าในการหาปริมาณวิเคราะห์ย่อมมีค่าคลาดเคลื่อนเสมอ อย่างไร ก็ตาม เพื่อให้มีการนำผลการวิเคราะห์ที่ได้มาใช้ได้ด้วยความมั่นใจ จึงมีความจำเป็นอย่างยิ่งที่ต้องนำระเบียบวิธีทางสถิติมาช่วยเพื่อตัดสินใจที่จะยอมรับ หรือปฏิเสธผลการวิเคราะห์ ดังนั้นสถิติจึงได้เข้ามามีบทบาทต่อเคมีวิเคราะห์อย่างมาก

ความหมายของคำที่ใช้ในสถิติ

1. ประชากร (population) หมายถึงมวลรวมของหน่วยทุกหน่วยที่เราสนใจศึกษาหรือหน่วยทุกหน่วยที่ให้ข้อมูลแก่เรา
2. ตัวอย่าง (sample) หมายถึงตัวแทนเพียงบางส่วนของสิ่งที่เราสนใจที่เลือกมาจากประชากร
3. พารามิเตอร์ (parameter) หมายถึงค่าต่าง ๆ ที่แสดงลักษณะของประชากร
4. ตัวสถิติ (statistics) หรือ ตัวประมาณ (estimator) หมายถึงค่าต่าง ๆที่แสดงลักษณะของตัวอย่าง

ตัวสถิติที่ใช้ในเคมีวิเคราะห์

1. มัชฌิมเลขคณิต (arithmetic mean) เรียกสั้น ๆ ว่า ค่าเฉลี่ย เป็นค่าที่ใช้วัดตำแหน่งส่วนกลางที่มีผู้นิยมใช้มากที่สุด เพราะสามารถสื่อความหมายหรือทำความเข้าใจได้ง่ายนอกจากนี้ยังมีคุณสมบัติที่ดีในทางสถิติเพื่อการอนุมานสถิติต่อไป ค่าเฉลี่ยคำนวณได้จากสูตร

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$\bar{x}$ = ค่าเฉลี่ยตัวอย่าง(sample mean)

n = จำนวนตัวอย่าง

2. ค่าแปรปรวน (variance) คือเฉลี่ยกำลังสองของผลต่างของค่าสังเกตแต่ละค่าจากค่าเฉลี่ย

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

ค่าแปรปรวนเป็นค่าที่สำหรับใช้วัดการกระจายของข้อมูล กล่าวคือ ดูว่าตัวแปร (X_i) แต่ละตัวมีความแตกต่างจากค่าเฉลี่ยมากน้อยเพียงใด ถ้าค่าแปรปรวนมีค่ามากแสดงว่าข้อมูลมีความแตกต่างกันมาก แต่ถ้าค่าแปรปรวนมีค่าน้อยแสดงว่าข้อมูลมีความแตกต่างกันน้อย โดยทั่วไปชุดข้อมูลทำการรวบรวมจากตัวอย่างหรือตัวแทนบางหน่วยของประชากรจึงเรียกว่า "ค่าแปรปรวนตัวอย่าง" (sample variance) ใช้สัญญลักษณ์ S^2

3. ค่าเบี่ยงเบนมาตรฐาน (standard deviation) คือรากที่สองของค่าแปรปรวน เป็นค่าที่ใช้วัดการกระจายของข้อมูลเช่นเดียวกับค่าแปรปรวน แต่เป็นที่นิยมใช้มากกว่า เนื่องจากค่าแปรปรวนมีหน่วยเป็นกำลังสอง

ถ้าค่าเบี่ยงเบนมาตรฐานมีค่ามากแสดงว่าข้อมูลมี ความแตกต่างกันมาก ถ้าค่าเบี่ยงเบนมาตรฐานมีค่าน้อยข้อมูลก็มีความแตกต่างกันน้อย ค่าเบี่ยงเบนมาตรฐานที่คำนวณจากข้อมูลทุกตัว เรียกว่าค่าเบี่ยงเบนมาตรฐานประชากร (population standard deviation) ใช้สัญญลักษณ์ว่า σ ค่าเบี่ยงเบนมาตรฐานที่คำนวณจากข้อมูลบางส่วนของประชากร เรียกว่าค่าเบี่ยงเบนมาตรฐานตัวอย่าง (sample standard deviation) ใช้สัญญลักษณ์ว่า S.D. หรือ S

$$S.D. = \sqrt{\sum \frac{(X_i - \bar{X})^2}{n - 1}}$$

4. สัมประสิทธิ์ความแปรผัน (coefficient of variation, CV) เป็นค่าสำหรับใช้วัดการกระจายของข้อมูล แต่คิดในรูปของการกระจายสัมพัทธ์ ไม่มีหน่วยและคำนวณออกมาเป็นเปอร์ เซนต์ จึงมีประโยชน์มากในการเปรียบเทียบข้อมูลที่มีหน่วยต่างกันหรือข้อมูลที่มีขนาดต่างกันมาก ถึงแม้จะมีหน่วยเดียวกันก็ตาม

$$\%\ CV = \frac{S.D.x100}{\bar{x}}$$

ค่าสถิติทั้ง 4 ตัวที่กล่าวมาแล้วข้างต้นใช้สำหรับการสรุปข้อมูลเชิงพรรณนา ซึ่งใช้เป็นประโยชน์สำหรับการวิเคราะห์และแปลผลสถิติอนุมานต่อไป ทั้งนี้เพื่อนำมาสู่การตัดสินใจสรุปผลหรืออ้างอิงไปสู่ประชากรที่สุ่มตัวอย่างมา

ตัวสถิติที่ใช้ทดสอบสมมติฐาน

1 กรณีเป็นข้อมูล 1 ชุด เป็นการทดสอบสมมติฐานเพื่อเปรียบเทียบค่าเฉลี่ยของข้อมูลชุดนั้นกับค่าที่แท้จริงซึ่งเป็นค่าที่ยอมรับ

ข้อสมมติเบื้องต้นค่าที่วัดหรือวิเคราะห์ได้จากแต่ละหน่วยตัวอย่างจะเป็นอิสระต่อกันและมีการแจกแจงแบบปกติ และประกอบด้วยหน่วยตัวอย่างน้อยกว่าหรือเท่ากับ 30 และไม่ทราบค่าแปรปรวน ตัวสถิติที่ใช้ทดสอบคือ

$$t = \frac{(x - \mu_0)}{(\frac{s}{\sqrt{n}})}$$

ตัวอย่างที่ 1 ในการวิเคราะห์ทดสอบปริมาณปรอท (SRM) [ซึ่งมีปริมาณอยู่ 38.9 %]

ด้วย Cold vapor atomic absorption ได้ปริมาณปรอทดังนี้ 38.9 %, 37.4 % และ 37.1% จากผลการวิเคราะห์ที่ได้ ต้องการทราบว่าค่าที่ได้มีความแตกต่างจากค่าจริงหรือไม่ที่ P=0.05

$$H_o \quad \bar{X} = \mu_0$$

$$H_a \quad \bar{X} \neq \mu_0$$

$$t = \frac{(x - \mu_0)}{\left(\frac{s}{\sqrt{n}}\right)}$$

μ_o = 38.9 %

$\bar{x}$ = 37.8%

s = 0.964 %

t_{cal} = 1.98

จากตารางการแจกแจงความถี่แบบ ที (t - distribution)

$t_{0.05,2}$ = 4.3

การตัดสินใจ ถ้า $t_{cal} < t_{0.05,2}$ ยอมรับ H_o กล่าวคือ $\bar{x} = \mu_o$

สรุปผล ในการวิเคราะห์ปริมาณปรอท (SRM) ด้วย cold vapor atomic absorption ผลการวิเคราะห์ไม่มีความแตกต่างจากค่าจริงอย่างมีนัยสำคัญที่ระดับความเชื่อมั่น 95 %

2. กรณีมีข้อมูล 2 ชุด ซึ่งข้อมูลทั้ง 2 ชุดนี้เป็นอิสระต่อกัน

ข้อสมมติเบื้องต้น ตัวแปรที่สนใจได้มาจากข้อมูล 2 ชุดที่มีการแจกแจงแบบปกติ (normal distribution) หรือใกล้เคียงปกติ และข้อมูลทั้ง 2 อาจมีค่าความแปรปรวนของตัวแปรที่สนใจเท่ากัน หรือไม่เท่ากันก็ได้

กรณีจำนวนตัวอย่างทั้ง 2 กลุ่ม น้อยกว่า 30 และไม่ทราบความแปรปรวนของข้อมูลทั้ง 2 กลุ่มและมีความแปรปรวนเท่ากัน**

ตัวสถิติที่ใช้ทดสอบคือ

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$Sp = \text{pooled variance} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

เมื่อ $n_1 + n_2 - 2$ คือ องศาความเป็นอิสระ (degree of freedom)

**** การที่จะนำ s_1^2 และ s_2^2 มาคำนวณเป็น s_p ต้องแน่ใจว่า ความแปรปรวนของข้อมูลทั้ง 2 กลุ่มนี้มีค่าเท่ากัน โดยทั่วไปสำหรับข้อมูลที่มีขนาด 10 ถึง 20 อัตราส่วนของความแปรปรวนของข้อมูลทั้ง 2 กลุ่ม ควรมีค่าไม่เกิน 3 หรือ 4 หากไม่มั่นใจควรทำการทดสอบความเท่ากันของค่าความแปรปรวนทั้ง 2 นี้โดยใช้ F test

ตัวอย่างที่ 2 จากการทดลองวิเคราะห์หาปริมาณโบรอนในพืชของ 2 วิธีซึ่งให้ผลการวิเคราะห์ดังนี้

	Spectrophotometric method	Fluorimetric method
จำนวนครั้งการทดลอง	10 ครั้ง	10 ครั้ง
ค่าเบี่ยงเบนมาตรฐาน(s)	0.3 μg/g	0.23 μg/g
ปริมาณโบรอนเฉลี่ย	28.8 μg/g	26.25 μg/g

ต้องการทราบว่าการวิเคราะห์ปริมาณโบรอนในพืชทั้ง 2 วิธีมีความแตกต่างกันหรือไม่ที่ระดับความเชื่อมั่น 95 %

สมมติฐาน $H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 \neq 0$

t_{cal} = 14.7

จากตารางการแจกแจงความถี่แบบ- ที (t - distribution)

$t_{0.05,18}$ = 2.1

การตัดสินใจ $t_{cal} > t_{0.05,18}$ ปฏิเสธ คือ $\mu_1 - \mu_2 \neq 0$

สรุปผล ผลจากการวิเคราะห์ปริมาณโบรอน ด้วย Spectrophotometric method กับ Fluorometric method มีความแตกต่างกันอย่างมีนัยสำคัญ ที่ระดับความเชื่อมั่น 95 %

ตัวอย่างที่ 3 จากการวิเคราะห์ปริมาณตะกั่วในอาหารชนิดหนึ่ง โดยการต้มด้วยกรดไฮโดรคลอริกโดยใช้เวลาต่างกัน ซึ่งให้ผลการทดลองดังนี้

เวลาที่ใช้ต้ม (นาที)	**ปริมาณตะกั่ว, mg/kg**
30	55 57 59 56 56 59
เวลาที่ใช้ต้ม (นาที)	**ปริมาณตะกั่ว, mg/kg**
75	57 55 58 59 59 59

ต้องการทราบว่าการวิเคราะห์ปริมาณตะกั่วโดยการต้มด้วยกรดในเวลา 30 นาที กับ 75 นาทีมีความแตกต่างกันหรือไม่ที่ระดับความเชื่อมั่น 95 %

สมมติฐาน $H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 \neq_o 0$

$t_{cal} = -0.84$

จากตารางการแจกแจงความถี่แบบ- ที (t - distribution)

$t_{0.05,10} = 2.23$

การตัดสินใจ $t_{cal} < t_{0.05,10}$ ยอมรับ H_o

สรุปผล จากการวิเคราะห์ปริมาณตะกั่วในอาหารชนิดหนึ่ง โดยการต้มด้วยกรดไฮโดรคลอริกด้วยเวลา 30 นาที กับ 75 นาที ไม่มีความแตกต่างกันอย่างมีนัยสำคัญ ที่ระดับความเชื่อมั่น 95 %

** กรณีมีข้อมูล 2 ชุด จำนวนตัวอย่างของข้อมูลทั้ง 2 กลุ่ม น้อยกว่า 30 และไม่ทราบความแปรปรวนของข้อมูลทั้ง 2 กลุ่ม **และมีความแปรปรวนไม่เท่ากัน** ตัวสถิติที่ใช้ทดสอบคือ

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

$$\text{องศาอิสระ} = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 + 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 + 1}} \quad \text{(โดยคิดเป็นจำนวนเต็ม)}$$

ตัวอย่างที่ 4 การวิเคราะห์ปริมาณ thiol ใน lysate ของเลือดจากอาสาสมัคร 2 กลุ่ม โดยที่กลุ่มแรกเป็นคนปกติ และกลุ่มที่2 เป็นกลุ่มที่ได้รับเชื้อรูมาตอยด์ ดังผลการทดลองดังนี้

thiol cocentration (mM)

กลุ่มคนปกติ	กลุ่มคนที่ได้รับเชื้อรูมาตอยด์
1.84	2.81
1.92	4.06
1.93	3.62
1.92	3.27
1.85	3.27
1.91	3.76
2.07	

$n_1 = 7$ $\quad n_2 = 6$

thiol เฉลี่ย $= 1.921$ mM $\quad = 3.465$ mM

$S_1^2 = 0.076$ mM $\quad S_2^2 = 0.440$ mM

ต้องการทราบว่าการวิเคราะห์ปริมาณ thiol ใน lysate ของเลือด จากอาสาสมัคร 2 กลุ่ม มีความแตกต่างกันหรือไม่ที่ระดับความเชื่อมั่น 95 %

สมมติฐาน $H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 \neq 0$

$t_{cal} = 8.5$

จากตารางการแจกแจงความถี่แบบ ที (t - distribution)

$t_{0.05,5} = 2.57$

การตัดสินใจ $t_{cal} > t_{0.05,5}$ ปฏิเสธ H_o

สรุปผล จากการวิเคราะห์ปริมาณ thiol ใน lysate ของเลือดจากอาสาสมัคร กลุ่มคนปกติ กับกลุ่มคนที่รับเชื้อรูมาตอยด์ มีความแตกต่างกันอย่างมีนัยสำคัญ ที่ระดับความเชื่อมั่น 95 %

การใช้ F- test เป็นสถิติที่ใช้เปรียบเทียบค่าความแปรปรวนของประชากร 2 กลุ่ม ซึ่งสมมติฐานที่ทดสอบมีทั้งแบบ one - tail test กับ two - tail test ทั้งนี้ขึ้นกับการตั้งสมมติฐาน ถ้า F เข้าใกล้ 1 จะยอมรับ H_o แต่ F มากกว่า หรือน้อยกว่า 1 จะปฏิเสธ H_o

* **กรณีเป็น one - tail test ทางขวา ตั้งสมมติฐาน ดังนี้**

ประชากรเดียว $H_o : \sigma^2 \leq \sigma_0^2$

$H_a \quad \sigma^2 > \sigma_0^2$

สองประชากร $H_o : \sigma_1^2 \leq \sigma_2^2$ หรือ

$H_o : \sigma_1^2 - \sigma_2^2 \leq 0$

$H_a \quad \sigma_1^2 \geq \sigma_2^2$ หรือ $H_a : \sigma_1^2 - \sigma_2^2 > 0$

จะใช้สมมติฐานนี้เมื่อต้องการทดสอบว่า มากกว่า หรือไม่มากกว่า

* **กรณีเป็น one - tail test ทางซ้าย ตั้งสมมติฐานดังนี้**

ประชากรเดียว $H_o : \sigma^2 \geq \sigma_0^2$

$H_a : \sigma^2 < \sigma_0^2$

สองประชากร $H_o : \sigma_1^2 \geq \sigma_2^2$ หรือ

$H_o : \sigma_1^2 - \sigma_2^2 \geq 0$

$H_a : \sigma_1^2 < \sigma_2^2$ หรือ $H_a : \sigma_1^2 - \sigma_2^2 < 0$

จะใช้สมมติฐานนี้เมื่อต้องการทดสอบว่า น้อยกว่า หรือ ไม่น้อยกว่า

กรณีเป็น two - tail test ตั้งสมมติฐาน ดังนี้

$H_o : \sigma_1^2 = \sigma_2^2$ หรือ $\sigma_1^2 - \sigma_2^2 = 0$

$H_a : \sigma_1^2 \neq \sigma_2^2$ หรือ $\sigma_1^2 - \sigma_2^2 \neq 0$

สถิติที่ใช้ทดสอบคือ $F = S_1^2 / S_2^2$

ตัวอย่างที่ 5 จากการวิเคราะห์หาปริมาณ COD ในห้องปฏิบัติการ โดยวิธี Proposed method กับ Standard method ได้ผลการทดลอง ดังนี้

	mean(mg/l)	Standard deviation (mg/l)	จำนวนครั้งการวิเคราะห์ (ครั้ง)
Standard method	72	3.31	8
Proposed method	72	1.51	8

$H_o: \sigma_1^2 \leq \sigma_2^2$

$H_a \quad \sigma_1^2 > \sigma_2^2$

$$F_{7,7} = \frac{(3.31)^2}{(1.51)^2}$$

$F_{cal\ 7,7} = 4.8$

$F_{cn\ 7,7} \quad 3.787$

การตัดสินใจ $F_{cal\ 7,7} > F_{cn\ 7,7}$ จึงปฏิเสธ H_o

สรุปผล การวิเคราะห์หาปริมาณ COD ด้วยวิธี Proposed method ไม่มีความแม่นยำมากกว่า Standard method

จากที่กล่าวมาแล้วข้างต้นจะเห็นได้ว่าสามารถนำวิธีทางสถิติเพื่อการพัฒนาคุณภาพของเคมีวิเคราะห์ได้โดยเฉพาะอย่างยิ่งเมื่อต้องการเปรียบเทียบผลการวิเคราะห์ระหว่างวิธีมาตรฐานกับวิธีที่พัฒนาขึ้นมาใหม่ หรือเปรียบเทียบผลการทดสอบที่วิเคราะห์ได้กับค่าที่แท้จริง หรืออาจทำการเปรียบเทียบความแปรปรวนระหว่างกลุ่มประชากรที่ศึกษา 2 กลุ่ม ว่ามีความแตกต่างกันหรือไม่ดังนั้น ผู้ที่จะนำวิธีทางสถิติมาใช้เป็นเครื่องมือในการพัฒนางานควรทราบว่า จะใช้สถิติตัวใดในการเปรียบเทียบ ทั้งนี้เพื่อกำหนดทิศทางของการทดสอบสมมติฐานให้เหมาะสม

เอกสารอ้างอิง

Miller ,J.C. and Miller,J.N. **Statistics for Analytical Chemistry** 2nd ed. , Chichester, W Sussex : Richard Clay , 1988

มัลลิกา บุนนาค **สถิติเพื่อการตัดสินใจ ภาควิชาสถิติ** กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย, 2539

วัลลภา ทาทอง การทดสอบสมมติฐานทางสถิติ **เอกสารประกอบการฝึกอบรมเรื่องการใช้สถิติ** และ **Control chart ในระบบคุณภาพห้องปฏิบัติการ** 2540 , มีนาคม 18 กรุงเทพมหานคร : กรมวิทยาศาสตร์บริการ,2540