# บทที่ 2
# การวิเคราะห์ข้อมูล
# (Data Analysis)

**การวิเคราะห์ข้อมูลจากการวิเคราะห์** การประเมินความถูกต้องของการวิเคราะห์เป็นสิ่งสำคัญ เนื่องจากต้องนำผลวิเคราะห์นี้แสดงการขึ้นทะเบียนอาหาร หรือฉลากอาหารสำหรับผู้บริโภคซึ่งผู้บริโภคสามารถตัดสินใจการซื้อผลิตภัณฑ์อาหารได้ถูกต้องตามความต้องการ หรือแนะนำผลิตภัณฑ์นี้ต่อผู้อื่นได้อย่างมั่นใจ ซึ่งการสรุปผลวิเคราะห์ข้อมูลควรประกอบด้วย เลขนัยสำคัญ ความเที่ยงตรง ความคลาดเคลื่อน เป็นต้น

## เลขนัยสำคัญ (Significant number)

ผลการวิเคราะห์ที่ได้จากการทดลองทุกครั้งจะมีหลายค่า หลังจากการคำนวณหาค่าเฉลี่ยเราต้องตอบให้มีเลขนัยสำคัญเพื่อระบุขอบเขตของความคลาดเคลื่อน โดยเลขนัยสำคัญ (significant figure) คือ จำนวนหลักของตัวเลขที่มีความหมายในผลการวิเคราะห์ เมื่อนับจำนวนเลขนัยสำคัญทำให้รู้ว่าหลักสุดท้ายเป็นค่าไม่แน่นอน(uncertainty) เช่นถ้าเราคำนวณค่าวิเคราะห์ได้ 8 กรัม แสดงว่าผลการวิเคราะห์อยู่ระหว่าง 7 กรัม กับ 9 กรัม ถ้าบอกค่าเป็น 8 $\pm$ 1 กรัม ในกรณีนี้เลขนัยสำคัญมี 1 หลัก (8) ที่มีค่าความไม่แน่นอนเท่ากับบวกหรือลบ 1

### แนวการนับเลขนัยสำคัญ

**1. เลขศูนย์ที่อยู่ระหว่างเลขสองตัวที่ไม่ใช่ศูนย์จะเป็นเลขนัยสำคัญเช่น**

708 มี เลขนัยสำคัญ 3 หลัก

708.1 มีเลขนัยสำคัญ 4 หลัก

**2. เลขหลักใดที่ไม่ใช่ศูนย์จะเป็นเลขนัยสำคัญเช่น**

567 กรัม มีเลขนัยสำคัญ 3 หลัก

239.56 กรัม มีเลขนัยสำคัญ 5 หลักเป็นต้น

**3. เลขศูนย์ที่อยู่ทางซ้ายของเลขที่ไม่ใช่ศูนย์ตัวแรกไม่ใช่เลขนัยสำคัญเพื่อแสดง**

ทศนิยม เช่น 0.02 กรัม มีเลขนัยสำคัญ 1 หลัก

0.0004289 กรัม มีเลขนัยสำคัญ 4 หลัก

**4. สำหรับเลขที่มีค่ามากกว่า 1 เลขศูนย์ที่อยู่ทางขวาของจุดทศนิยมจะเป็นเลขที่มีนัยสำคัญ** เช่น

5.0 กรัม มีเลขนัยสำคัญ 2 หลัก

20.058 กรัม มีเลขนัยสำคัญ 5 หลัก

**5. ถ้าเลขจำนวนใดลงท้ายด้วยเลขศูนย์ซึ่งไม่ได้อยู่ด้านขวาของจุดทศนิยม เลขศูนย์อาจไม่จำเป็นต้องเป็นเลขนัยสำคัญ** เช่น

180 กิโลกรัม อาจมีนัยสำคัญเป็น 2 หรือ 3

20,500 กิโลกรัม อาจมีนัยสำคัญเป็น 3 หรือ 4 หรือ 5

**6. การใช้เลขยกกำลังเพื่อความชัดเจนของเลขนัยสำคัญ เช่น** 20,400 กิโลกรัม สามารถใช้เป็นเลขยกกำลังดังนี้

$2.04 \times 10^4$ กิโลกรัม มีเลขนัยสำคัญ 3 หลัก

$2.040 \times 10^4$ กิโลกรัม มีเลขนัยสำคัญ 4 หลัก

$2.0400 \times 10^4$ กิโลกรัม มีเลขนัยสำคัญ 5 หลัก

**7. การนับเลข ถ้าต่ำกว่า 5 ปัดทิ้ง ถ้าหลักแรกหลังหลักที่นับนั้นมีค่าเท่ากันหรือมากกว่า 5 ให้ปัดขึ้นไปข้างหน้าเช่น**

5.628 ปัดเป็น 5.63

## การบวกและลบเลขนัยสำคัญ

1. การนับเลขนัยสำคัญของผลบวกหรือลบให้นับจากตัวตั้งที่มีเลขนัยสำคัญต่ำสุด**ทางขวา**ของจุดทศนิยมของตัวตั้งนั้น

$$\begin{array}{r} 68.912 \\ +\ 6.8 \\ \hline 75.712 \end{array} \longrightarrow \text{เลขนัยสำคัญ 1 หลักหลังจุดทศนิยม}$$

ผลลัพธ์คือ 75.7

$$\begin{array}{r} 5.246 \\ -\ 0.18 \\ \hline 5.066 \end{array} \longrightarrow \text{เลขนัยสำคัญ 2 หลัก}$$

ผลลัพธ์คือ 5.07

2. ในการคูณและการหาร กำหนดเลขนัยสำคัญของผลลัพธ์ จากเลขนัยสำคัญของตัวตั้งที่มีเลขนัยสำคัญน้อยที่สุด

เช่น 1.12 X 16.512 = 18.4934

ผลลัพธ์คือ 18.50

$$\frac{16.512}{1.143} = 14.4462$$

ผลลัพธ์คือ 14.45

**ตัวอย่าง 2.1** ตัวเลขต่อไปนี้มีจำนวนเลขนัยสำคัญ (significant number) แสดงในวงเล็บ

1. 10 ลิตร (1)
2. 11.3 ลิตร (3)
3. 0.020 กรัม (2)
4. 40,000 ลิตร (1)
5. 1.00 กรัม (3)
6. 0.00125 กรัม (3)
7. 12,105 กรัม (5)
8. 10.453 ลิตร (5)

### การฝึกคำนวนเลขนัยสำคัญ

1. 26.56 + 2.708 + 126.5 = ?
2. 218.14 – 14.715 = ?
3. 205.721 + 17.09712 + 0.78 + ?
4. 117- 0.56 + 4.106 = ?
5. 4.08 x 3.7 = ?
6. 0.0078 x 143 = ?
7. 0.435 x (80-23.8) = ?
8. $4.00 \times 10^4 - 1.2 \times 10^2$ = ?
9. 35 x 2.00 = ?
10. $(4.4)^4$ = ?

## ความแม่นและความละเอียดเที่ยงตรง (accuracy and precision)

**ความแม่น (accuracy)** คือผลการวิเคราะห์ที่มีค่าใกล้เคียงระหว่างค่าที่ทดสอบกับค่าจริง วิธีทดสอบค่า accuracy โดยทำการวิเคราะห์ Spiked samples อย่างน้อย 3 ความเข้มข้นๆละ

10 ซ้ำ โดยให้แต่ละความเข้มข้นครอบคลุมช่วงความเข้มข้นของการทดสอบ เช่นเราวิเคราะห์ความเข้มข้นของโปรตีนในตัวอย่างอาหารชนิดหนึ่งได้ 12.16% ส่วนค่าความถูกต้องเป็น 12.15% แสดงว่าการวิเคราะห์มีความแม่นยำสูง (high accuracy)accuracy แสดงในรูปของ % recovery

**ความเที่ยงตรง (precision)** คือผลการวิเคราะห์ซึ่งทำการทดลองหลายครั้งในสภาวะเดียวกันและได้ค่าใกล้เคียงกัน หรือคือความแม่นยำของการวัดซ้ำหลายๆครั้ง ซึ่งแสดงผลโดยค่า standard deviation ซึ่งก็คือ random error

**ตัวอย่าง 2.2** เช่น ในการทดลองหาปริมาณตะกั่วในปลาหมึก ซึ่งค่าที่ถูกต้องคือ 0.1028 มิลลิกรัม ได้ผลการทดลองดังนี้คือ

| | A | B | C |
|---|---|---|---|
| | 0.1001 | 0.1026 | 0.1004 |
| | 0.1003 | 0.1028 | 0.1998 |
| ค่าเฉลี่ย | **0.1002** | **0.1027** | **0.1501** |
| | มี Precision | มี Precision | ไม่มี Precision |
| | ไม่มีaccuracy | มี accuracy | ไม่มี accuracy |

**ความคลาดเคลื่อน (Error)** ความคลาดเคลื่อนจากค่าจริง ทำให้ทราบว่าผลการวิเคราะห์มี accuracy เพียงใด เช่นเดียวกับค่าเบี่ยงเบน (Deviation) ไปจากค่าเฉลี่ย ก็สามารถบอกว่าผลวิเคราะห์มี accuracy เพียงใด ความคลาดเคลื่อนแบ่งได้ 2 ชนิดคือ ความคลาดเคลื่อนแบบควบคุมได้ (Determinate Error) เป็นความคลาดเคลื่อนที่รู้ว่าคลาดเคลื่อนไปด้านใด สามารถควบคุมได้ และอีกชนิดหนึ่งคือความคลาดเคลื่อนที่แก้ไขไม่ได้ (Indeterminate error หรือ Random Error) ความคลาดเคลื่อนชนิดนี้เปลี่ยนแปลงได้ตลอด ไม่สามารถควบคุมได้

**ความคลาดเคลื่อนแบบควบคุมได้ (Determinate Error)** แบ่งได้เป็น 4 ชนิด คือ กระบวนการวิเคราะห์ ผู้วิเคราะห์ สารเคมี และเครื่องมือ

**1. ความคลาดเคลื่อนจากกระบวนการวิเคราะห์ (Method Error)** เกิดจากการวิเคราะห์ที่ไม่สามารถแยกสารประกอบออกจากสารตัวอย่างได้หมด ความคลาดเคลื่อนแบบนี้

สังเกตจากการเปลี่ยนแปลงสีของ Indicator วิธีแก้ไขอาจต้องทำ Blank เพิ่ม เพื่อหาปริมาณของสารเคมีที่ใช้ปริมาณ Indicator พอดีกับสารละลายที่ไม่มีตัวอย่างอยู่ เป็นการแก้ไขความคลาดเคลื่อนซึ่งเกิดจากจุดยุติและจุดสมมูลของปฏิกิริยาของการวิเคราะห์

**2. ความคลาดเคลื่อนจากผู้วิเคราะห์ (Personnel Error)** ผู้วิเคราะห์ควรมีความซื่อสัตย์ในการบันทึกผลการทดลอง ผู้ทดลองบางคนศึกษาผลการวิเคราะห์ให้เป็นไปในทางบวกแล้วค่อยบันทึก หรือเทคนิคที่นำมาวิเคราะห์ค่อนข้างซับซ้อน ผู้วิเคราะห์ไม่เข้าใจและทำผิดพลาด

**3. ความคลาดเคลื่อนจากสารเคมี** ควรเลือกใช้สารเคมีที่มีความบริสุทธิ์สูงจะลดความคลาดเคลื่อนในเรื่องนี้ได้ ถ้าเป็นงานวิจัยควรเลือกสารเคมีที่นักวิจัยอื่นๆนิยมใช้

**4. ความคลาดเคลื่อนจากเครื่องมือ** อาจเกิดจากมาตรวัดของเครื่องมือไม่ตรงกับความจริง ความสะอาดของเครื่องแก้ว วิธีแก้ไขโดยต้องมีการตรวจสอบความแม่นยำของเครื่องก่อนใช้ทุกครั้งสำหรับเครื่องมือที่สามารถตรวจสอบเองได้ แต่ถ้าการตรวจสอบต้องมาจากบริษัทที่จำหน่าย ก็ต้องตั้งงบประมาณมาตรวจเป็นครั้งคราว

**ความคลาดเคลื่อนแบบควบคุมไม่ได้** เกิดจากความผันแปรของวิธีวิเคราะห์ความคลาดเคลื่อนชนิดนี้มักเกิดขึ้นในจุดสูงสุดหรือต่ำสุดของมาตรวัดของเครื่องมือผู้วิเคราะห์ไม่สามารถบอกค่าที่วัดได้ชัดเจนถูกต้อง

**การทวนซ้ำของการวัด (Repeatability of results of measurements)**

คือ ความถูกต้องใกล้เคียงกันระหว่างผลของการวิเคราะห์ปริมาณที่ทำซ้ำกันหลายๆครั้งที่ต่อเนื่องกัน โดยทำภายใต้สภาวะเดียวกันทั้งหมด

**สภาวะทวนซ้ำประกอบด้วย**

- ระเบียบวิธีการวิเคราะห์แบบเดียวกัน
- ผู้ทดลองคนเดียวตลอดงานทดลอง
- เครื่องวิเคราะห์เครื่องเดียวกัน
- สถานที่ทดลองที่เดียวกัน

- การทวนซ้ำที่ไม่ทิ้งช่วงเวลามากนัก

**ความทำซ้ำของการวิเคราะห์ (Reproducibility of results of measurement)**

คือ ความถูกต้องใกล้เคียงกัน ระหว่างผลของการวิเคราะห์ปริมาณที่เดียวกันหลายๆ ครั้ง โดยการวิเคราะห์ครั้งหนึ่งๆอยู่ภายใต้การเปลี่ยนแปลงเงื่อนไขของการวิเคราะห์ สภาวะที่เปลี่ยนแปลงประกอบด้วย

- หลักการวิเคราะห์ - วิธีวิเคราะห์
- ผู้วิเคราะห์ - เครื่องมือ
- มาตรฐานการวัด - สถานที่
- สภาวะที่ใช้ - เวลา

**ความไม่แน่นอนของการวัด (uncertainty of measurement)**

◆ เป็นตัวแปรที่เกี่ยวเนื่องกับผลของการวิเคราะห์ ที่มีคุณลักษณะของการกระจายของค่าซึ่งครอบคลุมปริมาณที่วิเคราะห์

◆ ตัวแปรอาจจะเป็นค่าเบี่ยงเบนมาตรฐาน หรือครึ่งหนึ่งของความกว้างของความเชื่อมั่น

◆ ผลของการวิเคราะห์ ประกอบด้วยค่าประมาณที่ดีที่สุดสำหรับปริมาณที่วัด และองค์ประกอบของความไม่แน่นอน ซึ่งรวมผลกระทบเชิงระบบที่ทำให้เกิดการเบี่ยงเบน เช่น องค์ประกอบสำหรับค่าปรับแก้ และมาตรฐานอ้างอิง

◆ ค่าความไม่แน่นอน

**8. %Recovery** = ค่าที่วัดได้ X 100 / ค่าจริง

**9. Corrective factor (ตัวประกอบการปรับแก้)** คือ ตัวเลขที่ชดเชยสำหรับค่าผิดพลาดเชิงระบบ โดยนำมาบวกหรือลบกับค่าที่ยังไม่ปรับแก้เช่นเครื่องมือวัดค่าความหวานของผลไม้ที่อุณหภูมิหนึ่งแต่ผู้วิเคราะห์นำไปวัดอีกอุณหภูมิหนึ่ง ถ้าต้องการค่าที่ถูกต้องก็ต้องบวกค่าปรับแก้เข้าไป

## ความน่าเชื่อถือของการทดลอง

ผลการวิเคราะห์ที่ได้จากการทดลองแต่ละครั้งต้องมีความน่าเชื่อถือ ซึ่งจะทำให้ผู้วิเคราะห์เกิดความมั่นใจในการทดลอง ผลการทดลองสามารถพิสูจน์ได้หลายรูปแบบเช่น ค่า **Confident limit** เป็นวิธีที่นำค่าเบี่ยงเบนมาตรฐาน (s) มาใช้เราสามารถคำนวณผลการทดลองอยู่ในช่วงที่น่าเชื่อถือ กี่เปอร์เซ็นต์

โดยสูตร $\mu = \bar{x} \pm \frac{sf}{\sqrt{n}}$

$\mu$ คือค่าที่ถูกต้องที่ต้องการจากผลการทดลอง

$f$ เป็นตัวเลขจากตาราง

n จำนวนข้อมูล

ค่าที่ถูกต้อง ($\mu$) ควรอยู่ในขอบเขตของ $\bar{x} \pm \frac{sf}{\sqrt{n}}$

**ตัวอย่างที่ 2.3** ผลการทดลอง 5 จำนวน ค่า $\bar{x}$ = 35.25 ค่า S = 0.08 ที่ระดับความน่าเชื่อถือ 95% จงคำนวณค่าที่ถูกต้อง หาค่า $f$ จากตารางที่ระดับความน่าเชื่อถือ 95% มีค่าเป็น = 2.78

$$\mu = 3.25 \pm \frac{(2.78)(0.08)}{\sqrt{5}}$$

$$= 3.25 \pm \frac{0.2224}{2.236}$$

$$= 3.25 \pm 0.0995$$

35.35 $\longleftrightarrow$ 35.15

จากผลการคำนวณ 20 ครั้ง เราสามารถคาดคะเนว่า ผลการทดลองมีความถูกต้องอยู่ระหว่าง 35.35 และ35.15 จะมีเพียง 1 ครั้ง ส่วน 19 ครั้งที่อยู่นอกขอบเขตนี้

**ตารางที่ 2.1** แสดงค่า ระดับความน่าเชื่อถือ (*f* )

| จำนวนข้อมูล(n) | ระดับความน่าเชื่อถือ (confidence level) | | |
|---|---|---|---|
| | 90 % | 95 % | 99 % |
| 3 | 2.92 | 4.30 | 9.92 |
| 4 | 2.35 | 3.18 | 5.84 |
| 5 | 2.13 | 2.78 | 4.60 |
| 6 | 2.02 | 2.57 | 4.03 |
| 7 | 1.94 | 2.45 | 3.71 |
| 8 | 1.90 | 2.36 | 3.50 |
| 9 | 1.86 | 2.31 | 3.36 |
| 10 | 1.83 | 2.26 | 3.25 |
| 11 | 1.81 | 2.23 | 3.17 |
| 12 | 1.80 | 2.20 | 3.11 |
| 13 | 1.78 | 2.18 | 3.06 |
| 14 | 1.77 | 2.16 | 3.01 |
| 15 | 1.76 | 2.14 | 2.98 |
| 16 | 1.73 | 2.10 | 2.88 |

## การตัดข้อมูลทิ้ง (rejection of data)

เมื่อข้อมูลชุดหนึ่งประกอบด้วยค่าๆหนึ่งที่แตกต่างไปอย่างมากจากค่าอื่น ทำให้ต้องตัดสินใจว่าจะเก็บข้อมูลนั้นเพื่อการวิเคราะห์ต่อหรือตัดข้อมูลนั้นทิ้งไป ในทางสถิตินั้นสามารถใช้ คิว-เทสท์ (Q-test) เป็นเกณฑ์ในการตัดสินว่า จะตัดข้อมูลทิ้งหรือไม่ โดยเริ่มต้นจากการดึงข้อมูลที่ต้องสงสัยนั้นออก สมมติว่าเป็น $x_1$ ต่อจากนั้นเอาข้อมูลที่เหลือมาเรียงลำดับ จากค่ามากไปน้อยหรือน้อยไปมากก็ได้ แล้วจึงหาผลต่างระหว่างข้อมูลที่ต้องสงสัย กับข้อมูลที่มีค่าใกล้เคียงที่สุด ซึ่งจะเรียกว่า ความห่าง (spread) ตัวอย่างเช่น ถ้ามีข้อมูล 5 ค่าดังนี้ .

$x_1 \ldots\ldots x_2\ x_3\ x_4\ x_5$ ความห่างจะเท่ากับ $/ x_2 - x_1 /$

ต่อจากนั้นให้หาค่าพิสัย (range) โดยที่ พิสัยเท่ากับ $/ x_n - x_2 /$ เมื่อ n ซึ่งเป็นจำนวนข้อมูลมีค่าตั้งแต่ 3 ถึง 7 แต่ถ้าข้อมูลมีจำนวนเท่ากับ 8 , 9 หรือ 10 พิสัยจะเท่ากับ $/ x_{n-1} - x_1 /$ (ไม่ควรใช้ Q-test กับข้อมูลที่มีจำนวนมากกว่า 10) แล้วจึงคำนวณค่า Q ($0 < Q < 1$) ดังสมการ

$$Q_C = \frac{\text{ความห่าง}}{\text{พิสัย}} \qquad (2.14)$$

นำค่า Q ที่คำนวณได้นี้ ($Q_C$) ไปเปรียบเทียบกับค่า Q ในตารางที่ 2.3 ($Q_t$) ถ้า $Q_C < Q_t$ สามารถตัดข้อมูลที่ต้องสงสัยทิ้งได้

ในบางครั้ง อาจเป็นไปได้ที่ข้อมูลที่เหลือหลังจากการตัดข้อมูลที่ต้องสงสัยในครั้งแรกทิ้งไปแล้วอาจจะมีข้อมูลที่ต้องสงสัยว่าจะต้องตัดทิ้งอีกหรือไม่ในกรณีเช่นนี้ให้ดำเนินการเช่นเดียวกันกับที่ได้ทำมาแล้วกับข้อมูลที่ต้องสงสัยข้อมูลแรกแต่อย่างไรก็ตามไม่ควรใช้ Q-test มากกว่าสองครั้งกับข้อมูลชุดหนึ่ง ๆ

**การวิเคราะห์ข้อมูลที่จะตัดทิ้งมี 2 วิธีคือ**

**1. ใช้กฎ 4d rule**

**2. Method of Dixon's test หรือ Q test**

**1. ใช้กฎ 4d rule** วิธีนี้ใช้กับข้อมูลจำนวนน้อยซึ่งอย่างต่ำต้องมีข้อมมูล 4 ชุด

**วิธีคำนวน**

1.1 ต้องมีข้อมูลอย่างต่ำ 4 ชุด

1.2 หาค่าจุดกลาง (mean) ของข้อมูลที่ยอมรับ ข้อมูลที่สงสัยไม่ต้องมาคิด

1.3 คำนวนค่า standard deviation (SD)

1.4 คำนวนความแตกต่างของข้อมูลแต่ละชุดจาก mean

1.5 ถ้าค่าความแตกต่างเท่ากับหรือมากกว่า 4 เท่าของค่า mean ข้อมูลนั้นตัดทิ้งได้

**ตัวอย่าง 2.4** การวิเคราะห์ตัวอย่างอาหารชุดหนึ่งมีข้อมูลดังนี้ ขอให้คำนวนว่ามีข้อมูลใดควรตัดทิ้ง 10.15, 10.17, 10.34, 10.20

**ขั้นตอนที่ 1** เรียงข้อมูล

10.15

10.17

10.20

10.33 ข้อมูลนี้เป็นข้อมูลที่สงสัยควรตัดทิ้งหรือไม่

**ขั้นตอนที่ 2** หาค่าจุดกลาง (mean) ของข้อมูลชุดนี้

(10.15 + 10.17 + 10.20 ) /3 = 10.17

**ขั้นตอนที่ 3** หาความห่างของข้อมูลจาก mean (ไม่คิดเครื่องหมาย)

1 10.15 - 10.17 = 0.02

2 10.17 - 10.17 = 0.00

3 10.20 - 10.17 = 0.03

4 10.33 - 10.17 = 0.16

**ขั้นตอนที่ 4** หาค่าจุดกลาง (mean) ของข้อมูลที่ยอมรับ

(0.02 + 0.00 + 0.03)/3 = 0.017

**ขั้นตอนที่ 5** เปรียบเทียบค่าที่สงสัยว่ามากกว่าหรือน้อยกว่า 4 เท่า ของmeanในขั้นตอนที่ 4

0.017 x 4 = 0.07

0.16 > 0.07

ดังนั้นข้อมูลที่สงสัย คือ 10.33 ตัดทิ้งได้

**2. Method of Dixon's test หรือ Q test**

ใช้ทดสอบค่าสงสัยที่จะตัดทิ้งกับชุดข้อมูล 2-10 ชุด คำนวณ $Q_{cal}$ ได้จาก

$$Q_{cal} = \frac{X2 - X1}{Xn - X1}$$ **กรณีค่าสงสัยน้อยกว่าปกติ**

$$Q_{cal} = \frac{Xn - Xn - 1}{Xn - X1}$$ **กรณีค่าสงสัยมากกว่าปกติ**

$X_{1}, X_2, X_3, \ldots X_{n-1}, X_n$ เป็นข้อมูลที่ได้จากการวัดโดยเรียงลำดับจากน้อยไปมาก

$X_n - X_1$ คือ ค่าพิสัย (range)

นำค่า $Q_{cal}$ ที่คำนวณได้จากสมการข้างต้นไปเทียบกับ $Q_{crit}$ ในตารางแสดง Q

$Q_{cal} > Q_{crit}$ ค่าที่สงสัยตัดทิ้งได้

$Q_{cal} < Q_{crit}$ ค่าที่สงสัยตัดทิ้งไม่ได้

**ตารางที่ 2.2** ตารางแสดงค่า $Q_{crit}$ สำหรับการตัดข้อมูลทิ้ง

| จำนวนข้อมูล | องศาความอิสระ | $Q_{crit}$ (confidence limit) | | |
|---|---|---|---|---|
| (n) | (n - 1) | 90% | 96% | 99% |
| 3 | 2 | 0.94 | 0.98 | 0.99 |
| 4 | 3 | 0.76 | 0.85 | 0.93 |
| 5 | 4 | 0.64 | 0.73 | 0.82 |
| 6 | 5 | 0.56 | 0.64 | 0.74 |
| 7 | 6 | 0.51 | 0.59 | 0.68 |
| 8 | 7 | 0.47 | 0.54 | 0.63 |
| 9 | 8 | 0.44 | 0.51 | 0.60 |
| 10 | 9 | 0.41 | 0.48 | 0.57 |

**ตัวอย่างที่ 2.4** ในการวิเคราะห์ตัวอย่างสารตัวอย่างหนึ่ง พบว่ามีเปอร์เซ็นต์ของโซเดียมคลอไรด์ (NaCl) เท่ากับ 55.95 , 56.00 , 56.04 , 56.08 , และ 56.23 อยากทราบว่าค่าเปอร์เซนต์ NaCl ค่าสุดท้ายที่วิเคราะห์ได้นี้ สมควรจะตัดทิ้งหรือนำไปคิดค่าเฉลี่ย

**วิธีทำ**

ค่าความห่างของค่าที่สงสัย = 56.23 – 56.08 = 0.15%

ค่าพิสัย (เมื่อ n = 5) = 56.23 – 55.95 = 0.28%

ดังนั้น $Q_{cal} = \dfrac{0.15}{0.28} = 0.54$

จากตารางที่ 2.2 สำหรับข้อมูล 5 ชุด ( n= 5) ค่า $Q_{crit}$ = 0.64 ที่ 90% confidence limit
จะเห็นว่า $Q_{cal} < Q_t$
ดังนั้น ข้อมูล **56.23 จึงตัดทิ้งไม่ได้**

## การหาความลงตัวของข้อมูลโดยวิธีกำลังสองน้อยที่สุดเชิงเส้น

**(linear least square data fit)**

การวิเคราะห์ปริมาณสารด้วยวิธีเชิงอุปกรณ์ ทำให้ผู้ทำการวิเคราะห์จำเป็นต้องสร้างกราฟมาตรฐาน (calibration curve) ซึ่งมักจะเป็นเส้นตรง บ่อยครั้งที่การลากเส้นตรงอาจกระทำกันง่ายๆ โดยการทาบไม้บรรทัดกะให้เส้นตรงนั้นลากผ่านจุดข้อมูล โดยพยายามให้มีการกระจายของจุดเกิดอย่างดีที่สุดก็ตาม แต่วิธีที่ดีกว่าก็คือ การใช้สถิติช่วยในการหาเส้นตรงที่ดีที่สุด ที่จะเกิดความลงตัวกับจุดข้อมูลที่มี วิธีที่รู้จักกันดีคือ **วิธีกำลังสองน้อยที่สุด (least square method)**

ถ้าสมการของกราฟเส้นตรงเขียนได้เป็น

$$y = mx + b \qquad (2.15)$$

เมื่อ $y$ คือตัวแปรตาม และ $x$ เป็นตัวแปรอิสระ
$m$ คือค่าความชันของเส้นตรง
$b$ คือจุดตัดบนแกน $y$

โดยปกติ $y$ เป็นค่าที่อ่านได้จากเครื่องมือ ซึ่งเป็นฟังก์ชันค่า $x$ ที่เปลี่ยนไป ตัวอย่างเช่น ในกราฟมาตรฐานทาง สเปกโทรโฟโตเมตรี (spectrophotometric calibration curve) $y$ จะเป็นค่าการดูดกลืนแสง (absorbance) ในขณะที่ $x$ จะเป็นความเข้มข้นของสารมาตรฐาน

จุดประสงค์ของการใช้วิธีกำลังสองน้อยที่สุด คือ การหาค่าความชันและจุดตัดบนแกน $y$ ของข้อมูลที่มี ซึ่งสามารถหาได้จากสมการต่อไปนี้

$$m = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} = \frac{\overline{(xy)} - \overline{(x)}\overline{(y)}}{\sigma_x^2 (n-1)(n)} \quad (2.16)$$

$$b = \overline{y} - m\overline{x} \quad (2.17)$$

ค่าความเบี่ยงเบนมาตรฐานของค่าความชัน และ จุดตัดบนแกน y สามารถคำนวณได้จากความสัมพันธ์ดังนี้

$$\sigma_m = \frac{\sigma d}{\sqrt{\sigma x\, n-1}} \quad (2.18)$$

$$\sigma_b = \sigma m \sqrt{\frac{\sum x_i^2}{n}} \quad (2.19)$$

$$\sigma_d = \sqrt{\frac{\sum x_i^2 - b\sum y_i - m\sum x_i y_i}{n-2}} \quad (2.20)$$

ตัวอย่างต่อไปนี้ แสดงถึงการใช้วิธีกำลังสองน้อยที่สุด สำหรับผลการวิเคราะห์ในรูปแบบต่างๆ

**ตัวอย่างที่ 2.5** จากค่าการดูดกลืนแสงที่วัดได้ เมื่อใช้วิธีเฟลมอะตอมมิกแอบซอร์พชัน (flame atomic absorption) เพื่อวิเคราะห์ปริมาณ $Zn^{2+}$ ในสารมาตรฐานของ Zn ได้ผลดังนี้

| | | | | | |
|---|---|---|---|---|---|
| **X = ppm $Zn^{2+}$** | 0.50 | 1.00 | 1.50 | 2.00 | 2.50 |
| **Y= absorbance** | 0.130 | 0.200 | 0.350 | 0.430 | 0.490 |

จงเขียนสมการเส้นตรงจากการใช้วิธีกำลังสองน้อยที่สุดพร้อมกันนี้ให้ประมาณค่าความแม่นยำของความชันและจุดตัด ที่ความมั่นใจ 95 %

**วิธีทำ**

หาค่าเทอมต่างๆที่จะใช้ในสมการที่ (2.16) และ (2.17) จากข้อมูลข้างต้น ที่มี n = 5

$\sum x_i \quad = \; 5\bar{x} = 0.50+1.00+1.50+2.00+2.50 \; = \; 7.50$

$\sum y_i \quad = \; 5\bar{y} = 0.130+0.200+0.350+0.430+0.490 \; = \; 1.600$

$\sum x_i^2 \quad = \; (0.50)^2+(1.00)^2+(1.50)^2+(2.00)^2+(2.50)^2 \; = \; 13.750$

$\sum y_i^2 \quad = \; (0.130)^2+(0.200)^2+(0.350)^2+(0.430)^2+(0.490)^2 \; = \; 0.6.44$

$\sum x_i y_i = \; 5\overline{xy} \; = (0.50)(0.130)+(1.00)(0.200)+(1.50)(0.350)+(2.00)(0.430)+$

$(2.50)(0.490) \quad = \; 2.875$

$$m \quad = \quad \frac{5(2.875)-(7.50)(1.600)}{5(13.750)-(7.50)^2} \quad = \; 0.190$$

$$b \quad = \quad \frac{(1.600)}{5} - (0.190)\;\frac{(7.50)}{5} \quad = \; 0.0350$$

ดังนั้น สมการเส้นตรงที่ได้จะเป็น

$$y \quad = \quad 0.190X + 0.0350$$

ความมั่นใจของค่าจริงของ m และ b หาได้จาก

ช่วงความมั่นใจของค่าความชัน = $m \pm t_{CL,V}\,\sigma m$

ช่วงความมั่นใจของค่าความชัน = $b \pm t_{CL,V}\,\sigma b$

การหาค่า $\sigma_m$ และ $\sigma_b$ ต้องใช้สมการที่ (2.18) และ (2.19) แต่จำเป็นต้องหาค่า $\sigma_x$ และ $\sigma_b$ เสียก่อนจากสมการที่ (2.5) และ (2.20)

$$\sigma_x \quad = \quad \sqrt{\frac{13.750}{4} - \frac{(7.50)^2}{(5)(4)}} \quad = \; 0.7906$$

$$\sigma_d \quad = \quad \sqrt{\frac{(0.6044)-(0.0350)(1.600)-(0.190)(2.875)}{(5-2)}} \quad = \; 0.0268$$

นำค่า $\sigma_x$ และ $\sigma_d$ ที่ได้ไปหาค่า $\sigma_m$ แล้วจึงหาค่า $\sigma_b$

$$\sigma_m = \frac{0.0268}{0.7906\sqrt{4}} = 0.0169$$

$$\sigma_b = 0.0169\sqrt{\frac{13.750}{5}} = 0.028$$

$$t_{CL,V} = t_{95,n-2} = t_{95,3} = 3.182$$

ดังนั้น ค่าประมาณความแม่นยำของ

ความชัน $= 0.190 \pm (3.182)(0.0169) = 0.190 \pm 0.054$

จุดตัด $= 0.035 \pm (3.182)(0.028) = 0.035 \pm 0.089$

จากการหาความลงตัวของข้อมูลโดยวิธีกำลังสองน้อยที่สุด ซึ่งทำให้ได้สมการของกราฟเส้นตรงที่ดีที่สุด ที่สามารถนำไปใช้คำนวณหาค่าตัวแปร x ที่ไม่ทราบค่า ($x_u$) เมื่อทำการวัดค่า $y_u$ จากการทดลองได้

สมมติว่ามีข้อมูลจากการทดลองจำนวนเท่ากับ n ($x_i$ และ $y_i$) ที่นำมาหาความลงตัวด้วยวิธีกำลังสองน้อยที่สุด สำหรับสารตัวอย่างที่มีค่าตัวแปรอิสระเป็น $y_u$ ดังนั้นค่า $x_u$ ที่จะคำนวณได้จากสมการเส้นตรงจะเป็น

$$x_u = \frac{\bar{y}_u - b}{m} \qquad (2.21)$$

โดยมีค่าเบี่ยงเบนมาตรฐานของ $x_u$ ที่คำนวณได้จากสมการ

$$\sigma x_u = \frac{\sigma d}{m}\left[\frac{1}{p} + \frac{1}{n} + \frac{(x_u - \bar{x})^2}{\sigma x^2\,(n-1)}\right]^{1/2} \qquad (2.22)$$

ถ้า $\bar{y}_u$ เป็นค่าทางทฤษฎี (ไม่ใช้ได้จากการวัด) จะถือว่า p ในสมการที่ (2.22) จะมีค่าเป็น $\infty$

**ตัวอย่างที่ 2.6** สารละลายตัวอย่างของ $Zn^{2+}$ ที่ไม่ทราบความเข้มข้นจำนวนหนึ่ง เมื่อนำมาแบ่งเป็นสามส่วน แล้วนำแต่ละส่วนไปวิเคราะห์ โดยวิธีตามตัวอย่างที่ 2.14 พบว่า ค่าการดูดกลืนแสงเฉลี่ยของสารตัวอย่างทั้งสามส่วนมีค่าเท่ากับ 0.250 จงคำนวณหา

ก. ความเข้มข้นของ Zn ในสารละลายตัวอย่าง

ข. ความแม่นยำของความเข้มข้นที่คำนวณได้ ภายใต้ความมั่นใจ 95 %

**วิธีทำ**

ก. สมการเส้นตรง

$$y = 0.190x + 0.0350$$

แทนค่าจากสมการที่ (2.21)

$$\bar{y} = 0.250$$

$$x_u = \frac{0.250 - 0.0350}{0.190} = 1.10 \text{ ppm } Zn^{2+}$$

ข. ผลลัพธ์จากตัวอย่างที่ 2.14 และ ใช้สมการที่ (2.22)

$$\sigma x_u = \frac{0.0268}{0.190}\left[\frac{1}{3}+\frac{1}{5}+\frac{(1.13 - 1.50)^2}{(0.7906)^2 (4)}\right]^{1/2}$$

$$= 0.108 \text{ ppm } Zn^{2+}$$

ค่าความแม่นยำของ $x_u$ ที่ความมั่นใจ 95 %หาได้จาก $x_u \pm t_{CL,V}\, \sigma x_u$ เมื่อ V = n – 2 ดังนั้น

$$x_u = 1.13 \pm (t_{95,3})(0.108)$$

จากตารางที่ 2.1 $t_{95,3} = 3.182$ เพราะฉะนั้น

$$x = 1.13 \pm (3.182)(0.108)$$

$$= 1.13 \pm 0.34 \text{ ppm } Zn^{2+}$$

นอกจากการทำ least square fit ของข้อมูลที่ได้จากการทดลองดังตัวอย่างที่กล่าวมาแล้วนั้น โดยปกติแล้วผู้ทำการวิเคราะห์มักต้องการที่จะดูความสัมพันธ์ระหว่างตัวแปรทั้งสอง ($x_i$ และ $y_i$ ) มีความสัมพันธ์ระหว่างกันมากน้อยเพียงใด ซึ่งสามารถบอกได้ โดยการคำนวณค่า สัมประสิทธิ์สหสัมพันธ์ (correlation coefficient, r) ของข้อมูลจากสมการต่อไปนี้

$$r = \frac{\sum x_i y_i - n(\overline{xy})}{\sqrt{(\sum x_i^2 - n\bar{x}_i^2)(\sum y_i^2 - n\bar{y}_i^2)}} \qquad (2.23)$$

หรือ

$$r = \frac{n(\sum x_i y_i) - \sum x_i \sum y_i}{\sqrt{[n(\sum x_i) - (\sum x_i)^2][n(\sum y_i) - (\sum y_i)^2]}} \qquad (2.24)$$

ค่า r ที่คำนวณได้ จะมีค่าสูงสุดเท่ากับ 1 ซึ่งหมายถึงตัวแปรทั้งสองมีความสัมพันธ์กันอย่างเต็มที่ ในขณะที่ $r = 0$ หมายความว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กันเลย และถ้า $r = -1$ ซึ่งถือว่าเป็นค่าน้อยที่สุด จะหมายความว่า ความสัมพันธ์ของตัวแปรทั้งสองดังกล่าวกลับกันกับที่เขียนไว้เช่น แทนที่ y จะขึ้นกับ x เมื่อ x เป็นตัวแปรอิสระ แล้ว y เป็นตัวแปรตาม กลับกลายเป็น x เป็นตัวแปรตาม ขณะที่ y เป็นตัวแปรอิสระแทน

สำหรับการสร้างกราฟมาตรฐาน โดยใช้ least square fit นั้น ค่าสัมประสิทธิ์สหสัมพันธ์ ที่คำนวณได้สามารถสรุปได้ว่า กราฟเส้นตรงที่ได้นั้นมีลักษณะที่

| | | |
|---|---|---|
| พอใช้ได้ | ถ้า | $0.90 < r < 0.95$ |
| ดี | ถ้า | $0.95 < r < 0.99$ |
| ดีมาก | ถ้า | $r < 0.99$ |

แต่ถ้าการวัดค่าต่างๆ และการทดลองทำด้วยความระมัดระวังอย่างมาก อาจทำให้ค่า $r < 0.999$

**ตัวอย่างที่ 2.7** จากข้อมูลในตัวอย่างที่ 2.14 จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทั้งสอง

**วิธีทำ**

จากสมการที่ (2.24 ) และค่าต่าง ๆจากตัวอย่างที่ 2.14

$n = 5, \sum x_i = 7.50, \sum y_i = 1.600, \sum x_i \sum y_i = 2.875$

$\sum x_i^2 = 13.75, \sum y_i^2 = 13.75$

ดังนั้น

$$r = \frac{5(2.875) - (7.50)(1.600)}{\sqrt{[5(13.750) - (7.50)^2][5(0.6044) - (1.600)^2]}} = 0.99$$

$\therefore$ ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทั้งสอง = 0.99