

# การทดสอบการแจกแจงปกติ

อนุสิทธิ์ สุขม่วง

**การ**ทดสอบสมมติฐานต่างๆ ที่ใช้กันในปัจจุบันจะตั้งอยู่บนสมมติฐานของข้อมูลที่เป็นตัวแปรสุ่มและมีการแจกแจงปกติ นอกจากนั้นในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรที่มากกว่าสองกลุ่ม ยังสมมติฐานว่า ความแปรปรวนเท่ากัน ซึ่งการทดสอบแบบนี้เป็นการทดสอบที่ใช้พารามิเตอร์ (Parametric Test) ในบางสถานการณ์ ผู้ทดสอบไม่แน่ใจเกี่ยวกับลักษณะของประชากรที่สนใจ การทดสอบที่ใช้พารามิเตอร์ อาจทำให้ได้ผลที่คลาดเคลื่อนได้ เนื่องจากข้อมูลไม่เป็นไปตามสมมติฐานที่ตั้งไว้ ดังนั้นการทดสอบทางสถิติเพื่อการตัดสินใจ จึงควรใช้วิธีการทดสอบที่เรียกว่าการทดสอบที่ไม่ใช้พารามิเตอร์ (Nonparametric Test)

การทดสอบการแจกแจงปกติ มีวิธีทดสอบที่ใช้กันอยู่ คือ การทดสอบไคสแควร์ (Chi-squared test) แต่การทดสอบนี้ใช้ได้กับข้อมูลที่มีจำนวนมาก ปกติการทดสอบทางวิทยาศาสตร์จะไม่มีข้อมูลมากนัก Kolmogorov-Smirnov test จึงเป็นวิธีทดสอบที่เหมาะสมกับการทดสอบที่มีจำนวนข้อมูลไม่มาก

วิธีการทดสอบโดยวิธี Kolmogorov-Smirnov test ใช้เปรียบเทียบค่าของการแจกแจงความถี่สะสมของข้อมูลที่ทดสอบ (Empirical Distribution) กับค่าของการแจกแจงที่ต้องการทดสอบ คือ การแจกแจงปกติ ค่าสถิติ ค่าที่มากที่สุดของค่าความแตกต่างของค่าทั้งสอง การทดสอบทำได้ได้สองลักษณะ ลักษณะแรกทดสอบว่าการแจกแจงนั้นเป็นการแจกแจงปกติที่สงสัยหรือไม่ โดยที่ทราบค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน ลักษณะที่สองทดสอบว่าเป็นการแจกแจงปกติหรือไม่

ข้อมูลที่ใช้ทดสอบเป็นข้อมูลเชิงปริมาณ มีระดับมาตรการวัด มาตรฐานอัตราส่วน (Ratio scale) คือ การให้สัญลักษณ์ของข้อมูลที่เป็นตัวเลข แสดงปริมาณสามารถนำมาบวก ลบ คูณ หรือหาร กันได้

## ตั้งสมมติฐาน

สมมติฐานหลัก  $H_0$  : ตัวแปรมีการแจกแจงปกติ

สมมติฐานรอง  $H_1$  : ตัวแปรไม่มีการแจกแจงปกติ

## วิธีทดสอบ

1. เรียงข้อมูล จำนวน  $n$  ข้อมูลจากน้อยไปมาก ให้  $k$  เป็นอันดับของข้อมูลที่เรียงแล้ว

2. แปลงข้อมูลแต่ละค่าให้เป็นคะแนน  $Z$  โดยใช้สูตร

$$Z = \frac{x - \mu}{\sigma}$$

3. หาค่าความถี่สะสม  $F(z)$  ของการแจกแจงปกติจากตาราง  $Z$

4. หาค่าความถี่สะสมที่ได้จากการทดลอง ECDF(z)

$$ECDF(z) = \frac{1}{n} \sum_{i=1}^n$$

5. เปรียบเทียบ หาค่าความแตกต่างสูงสุดของค่าทั้งสอง ให้เป็น  $D_{\max}$

$$D_{\max} = |F(z) - ECDF(z)|$$

6. เปรียบกับค่าในตาราง The Kolmogorov test for normality

7. สรุปผล ยอมรับสมมติฐาน ถ้า  $D_{\max}$  น้อยกว่าค่าจากตาราง

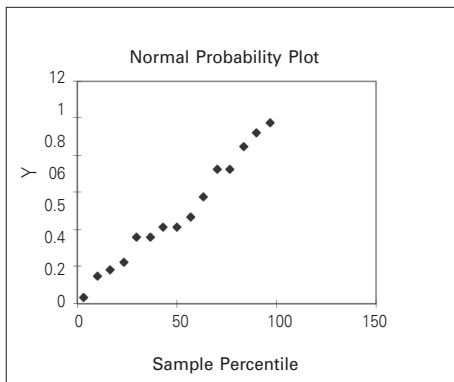
**ตัวอย่าง** ผลการวัดโดย Spectrophotometer

15 ครั้ง ได้ผลดังนี้ .3410 .3350 .3470 .3590 .3530 .3460 .3470 .3460 .3430 .3420 .3560 .3500 .3630 .3530 และ .3480 จงทดสอบว่า

1. ข้อมูลมีการแจกแจงปกติ

2. ข้อมูลมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0.35 ml. และส่วนเบี่ยงเบนมาตรฐาน 0.005 ml.

การตรวจสอบอย่างง่าย โดยใช้ Normal Probability Plot ใช้เครื่องมือใน Data analysis ของ MS Excel ได้กราฟ ดังนี้



กราฟมีรูปร่างคล้ายตัวอักษร S จึงสรุปได้ว่าข้อมูลมีการแจกแจงปกติ

## 1. การทดสอบ ข้อมูลมาจากการแจกแจงปกติ

(1) เรียงข้อมูลจำนวน  $n$  ข้อมูลจากน้อยไปมาก ให้  $k$  เป็นอันดับของข้อมูลที่เรียงแล้ว

(2) แปลงข้อมูลให้เป็นคะแนน  $Z$  โดยใช้สูตร

$$Z = \frac{x - \mu}{\sigma}$$

(3) หาค่า  $F(z)$  จากตารางค่า  $Z$  ในที่นี้จะใช้ฟังก์ชัน = NORMSDIST( )

(4) หาค่า ความถี่สะสมจากผลการทดลอง  $ECDF(z) = k * (1/n)$

(5) เปรียบเทียบค่าความแตกต่างของ  $F(z)$  และ  $ECDF(z)$  แต่ละค่า

### ดูผลจากตาราง

Point	x	z	F(z)	ECDF(z)	D =  F(z)-ECDF(z)
1	0.335	-1.8611	0.0314	0.0667	0.035301
2	0.341	-1.0400	0.1492	0.1333	0.015828
3	0.342	-0.9032	0.1832	0.2000	0.016791
4	0.343	-0.7663	0.2217	0.2667	0.044938
5	0.346	-0.3558	0.3610	0.3333	0.027649
6	0.346	-0.3558	0.3610	0.4000	0.039021
7	0.347	-0.2190	0.4133	0.4667	0.053343
8	0.347	-0.2190	0.4133	0.5334	0.120013
9	0.348	-0.0821	0.4673	0.6000	0.132746
10	0.35	0.1916	0.5760	0.6667	0.090731
11	0.353	0.6021	0.7265	0.7334	0.006914
12	0.353	0.6021	0.7265	0.8000	0.073584
13	0.356	1.0127	0.8444	0.8667	0.022319
14	0.359	1.4232	0.9227	0.9334	0.010717
15	0.363	1.9706	0.9756	1.0000	0.024433

$$D_{\max} = 0.1327$$

(6) หาค่าความแตกต่างสูงสุด  $D_{\max} = 0.1327$

(7) เปรียบเทียบค่าจากตาราง The Kolmogorov test for normality เมื่อ  $n=15$  เท่ากับ 0.219 ซึ่ง  $D_{\max}$  มีค่าน้อยกว่า

(8) สรุป ข้อมูลมีการแจกแจงปกติ

**2. การทดสอบ ข้อมูลจากการแจกแจงปกติ** ที่มีค่าเฉลี่ย 0.35 และส่วนเบี่ยงเบนมาตรฐาน 0.005 หรือไม่จากการทดสอบโดยวิธีเดียวกันโดยกำหนดค่าเฉลี่ย 0.35 และส่วนเบี่ยงเบนมาตรฐาน 0.005 จะได้ค่า  $D_{\max} = 0.259$  มากกว่าค่าจากตาราง The Kolmogorov test for normality เมื่อ  $n=15$  ซึ่งได้ค่าเท่ากับ 0.219 จึงไม่ยอมรับสมมติฐาน สรุป ข้อมูลไม่ได้มาจากการแจกแจงปกติ ที่มีค่าเฉลี่ย 0.35 และส่วนเบี่ยงเบนมาตรฐาน 0.005

ความรู้ที่ได้รับสามารถนำไปใช้ในการวิเคราะห์ผลการทดสอบ ซึ่งการทดสอบทางวิทยาศาสตร์มีข้อจำกัดเกี่ยวกับจำนวนตัวอย่าง มักจะใช้จำนวนตัวอย่างไม่มาก การทำให้ผลการวิเคราะห์เชื่อถือได้ต้องอาศัยหลักการทางสถิติอ้างอิง

## เอกสารอ้างอิง

- Kolmogorov-Smirnov Test. [Online] [Cite dated 4 June 2550] Available to internet : <http://www.physics.csbsju.edu/stats/KS-test.html>
- Mathwave Technologies. [Online] [Cite dated 4 June 2550] Available to internet : [http://www.mathwave.com/articles/goodness\\_of\\_fit.html](http://www.mathwave.com/articles/goodness_of_fit.html)
- Miller, James N. & Miller, Jane C. **Statistics and chemometrics for analytical chemistry**. 4<sup>th</sup> ed., New York : Prentice Hall, 1999.
- Otto, Matthias. **Chemometrics**. Weinheim, WILEY-VCH, 1998.
- Wu, Gaw Tzu, Twomey, Stanley L; and Thlers, Ralph E. Statistical evaluation of method comparison data. **Clinical Chemistry**, 1975, Vol. 21, No. 3, p 315-320.
- จุฬาลงกรณ์มหาวิทยาลัย. คณะวิทยาศาสตร์. ภาควิชาคณิตศาสตร์. **ความน่าจะเป็นและสถิติ**. พิมพ์ครั้งที่ 9. กรุงเทพมหานคร : สำนักพิมพ์ฟิสิกส์เซ็นเตอร์, 2544.