

# การคำนวณค่าทางสถิติสำหรับงานวิเคราะห์ทดสอบ

โดย นางสาวสุทธินี มีสุข  
ผู้ช่วยเจ้าหน้าที่ห้องปฏิบัติการ  
กลุ่มงานวิเคราะห์คุณภาพสิ่งแวดล้อม  
สำนักงานสิ่งแวดล้อมภาคที่ 13 (ชลบุรี)

สถิติ เป็นเครื่องมือที่มีคุณภาพมากที่ใช้ในการควบคุมคุณภาพ มีส่วนสำคัญในกระบวนการทำงานที่เกี่ยวข้องกับตัวเลขเป็นจำนวนมาก สามารถใช้ในการนำเสนอและเป็นข้อมูลเพื่อการตัดสินใจ และการปฏิบัติการที่ถูกต้อง เทคนิคทางสถิติเป็นเครื่องมือที่ใช้ในการตรวจหา แยกแยะ และลดความผิดพลาดของกระบวนการทำงาน โดยลดปัจจัยที่ทำให้เกิดความแปรผันซึ่งผลที่ได้ทำให้ผลวิเคราะห์ทดสอบเชื่อถือได้ มีความถูกต้อง และมีความเที่ยงซึ่งเป็นเป้าหมายสูงสุดของการวิเคราะห์ทดสอบ เพื่อให้ง่ายแก่การใช้งานโปรแกรม Excel เป็นโปรแกรมที่ง่าย สะดวกและสามารถใช้ในการคำนวณค่าทางสถิติได้โดยไม่ต้องใช้โปรแกรมสำเร็จรูปเช่น SPSS เพราะฉะนั้นจึงเป็นเรื่องจำเป็นที่ต้องรู้พื้นฐานของการใช้โปรแกรม Excel

## 1. พื้นฐานการใช้ Excel

1.1 การเขียนข้อความ ถ้าเป็นข้อความที่เป็นตัวอักษรสามารถเขียนข้อความลงไปได้ทันที แต่ถ้าต้องการเขียนเป็นตัวเลขแต่เป็นลักษณะของข้อความเช่นเลขที่บ้าน ให้ใช้เครื่องหมาย “'” นำหน้าข้อความที่ต้องการเขียน ถ้าเป็นสูตรหรือฟังก์ชัน ให้เขียนนำหน้าด้วยเครื่องหมาย “=” และถ้าต้องการรู้ว่าผลของการคำนวณนั้นๆ มีที่มาจากสูตรหรือฟังก์ชันใด ให้ใช้เครื่องหมาย “” นำหน้าสูตรนั้นๆ จากนั้นกด enter จะแสดงที่มาของผลลัพธ์นั้นๆ ได้

1.2 การ copy ข้อมูล มีด้วยกัน 4 แบบคือ

1.2.1 copy ทั่วไป คือใช้คำสั่ง copy & paste

1.2.2 copy แบบพิเศษ (เฉพาะ format) หรือ copy ลักษณะของตัวอักษรหรือรูปแบบต่างๆ จาก cell ต้นแบบ โดยหลังจากที่มี cell ที่ต้องการแล้ว ในช่อง cell ต้นแบบ เลือก tool bar รูปแปรงทาสี จากนั้น paste ไปที่ cell ที่ต้องการ

1.2.3 copy แบบ Auto Fill โดยการใช้ mouse คลิกไปที่มุมล่างขวามือของ cell ต้นแบบและลาก mouse ไปยังทิศทางที่ต้องการ ผลลัพธ์ที่ได้จะเป็นค่าที่ต่อเนื่อง การกำหนดค่า

ต้นแบบใช้งานเองโดยไปที่ Tool/Option/Custom List/กรอกค่าต้นแบบและเว้นแต่ละค่าด้วย เครื่องหมาย “,”/กด Add

1.2.4 copy worksheet ให้อยู่ในรูป picture โดยการเลือก cell ต้นแบบ/กด shift ค้าง/Edit/Copy Picture/เลือกตำแหน่งที่วาง/Paste Special

1.3 วิธีการเขียนสูตร ไปที่ Insert/Object/Microsoft Equation 3.0 จะมีรูปแบบและ ฟังก์ชันของการเขียนสูตรที่ถูกต้อง สะดวก สามารถนำมาใช้เขียนสำหรับจัดทำเอกสารคู่มือได้ (ใน โปรแกรม Microsoft Word สามารถทำได้เช่นกัน โดยเข้าไปที่เมนูเดียวกันนี้) หรือ ไปที่ Tool/Customize/ Command/ Insert/ Equation Editor แล้วลากไปไว้ที่ Toolbar

1.4 รูปแบบการอ้างอิง cell มี 3 ลักษณะ ได้แก่

1.4.1 การอ้างอิงแบบสัมพัทธ์ คือเมื่อมีการเปลี่ยนแปลงเช่นย้ายหรือคัดลอกสูตร จะ ทำให้ตำแหน่งของการอ้างอิงเปลี่ยนตามไปด้วย

1.4.2 การอ้างอิงแบบสัมบูรณ์ คือการล็อคตำแหน่งของ cell ที่ถูกอ้างอิง เมื่อมีการ ย้ายหรือคัดลอกสูตร cell อ้างอิงจะไม่เปลี่ยนตามไปด้วย โดยใช้เครื่องหมาย “\$” วางไว้ที่หน้า cell อ้างอิง (วางหน้าตัวอักษรและตัวเลข หรือใช้ปุ่มลัดโดยกดปุ่ม F4)

1.4.3 การอ้างอิงแบบผสม คือนำการอ้างอิงแบบสัมพัทธ์และแบบสัมบูรณ์มาใช้ใน สูตรเดียวกัน ขึ้นกับความต้องการในการใช้งาน

1.5 การใช้ฟังก์ชัน

1.5.1 ฟังก์ชันทางคณิตศาสตร์ เช่น สำหรับการรวมผลรวมคือ SUM(), จำนวนเต็ม และปัดเศษทศนิยมคือ INT(), อื่นๆ ทางคณิตศาสตร์ เช่น POWER(), SQRT() เป็นต้น

1.5.2 ฟังก์ชันทางสถิติ เช่น การจัดลำดับข้อมูลคือ MAX(), MIN(), QUANTILE() การ วิเคราะห์เชิงเส้น คือ INTERCEPT() การวัดการกระจายของข้อมูล คือ STDEV() การนับจำนวนคือ COUNT() เป็นต้น

## 2. สถิติเชิงพรรณนา

คือการแสดงข้อมูลของสถิติพื้นฐาน เช่น การวัดค่ากลาง การวัดการกระจาย การแสดงข้อมูล ด้วยกราฟ เป็นต้น

2.1 การวัดค่ากลาง เป็นการพยายามหาค่ากลางของข้อมูลชุดใดชุดหนึ่ง เพื่อใช้เป็นตัวแทน ของข้อมูลชุดนั้นๆ โดยค่ากลางที่นิยมใช้ ได้แก่ ค่าเฉลี่ยเลขคณิต จากสูตร  $\bar{x} = \frac{\sum x_i}{n}$

## 2.2 การวัดการกระจาย ที่นิยมใช้ได้แก่

2.2.1 พิสัย (Range, R) เป็นค่าแสดงการกระจายข้อมูล บอกช่วงการกระจายทั้งหมด 100% เพื่อให้ทราบว่ามียุคสูงสุด และค่าต่ำสุดเท่าใด จากสูตร  $R = X_{\max} - X_{\min}$

2.2.2 ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation, S) คือค่าเฉลี่ยของส่วนเบี่ยงเบนระหว่างข้อมูลแต่ละจำนวน กับค่าเฉลี่ยเลขคณิตของข้อมูลนั้น หาได้จากสูตร

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

2.2.3 ความแปรปรวน (Variance,  $S^2$ ) เป็นการยกกำลังสอง ของค่าส่วนเบี่ยงเบนมาตรฐาน ใช้ในการเปรียบเทียบการกระจายของข้อมูล 2 ชุด มีประโยชน์ในการรวมค่าความแปรปรวน

2.2.4 สัมประสิทธิ์การกระจาย ใช้เปรียบเทียบการกระจายของข้อมูลต่างชุดกัน เช่นมีค่าเฉลี่ยต่างกัน หรือส่วนเบี่ยงเบนมาตรฐานที่ต่างกัน หาได้จากสูตร  $\%RSD = \frac{S}{x} \times 100$

## 2.3 การแสดงข้อมูลด้วยกราฟ กราฟที่นิยมใช้แสดงข้อมูล ได้แก่

2.3.1 Histogram เป็นกราฟแท่งที่ใช้แสดงข้อมูลในลักษณะของความถี่ แกนนอนแสดงค่าของข้อมูลและแนวตั้งแสดงค่าของความถี่ของข้อมูลในช่วงต่างๆ

2.3.2 Dot plot เป็นการแสดงข้อมูลแบบต่อเนื่อง คล้ายๆ Histogram แต่ละค่าของข้อมูลที่เท่ากันจะแสดงเป็นจุดวางเรียงกันในแนวตั้ง

2.3.3 Stem and Leaf Plot (แผนภาพลำต้นและใบ) เป็นกราฟที่ใช้จำนวนตัวเลขแสดงเป็นกราฟ จะใช้หลักที่มีจำนวนสูงสุดเป็นแกนในแนวคอลัมน์เรียกว่า Stem และหลักของตัวเลขถัดไปจะวางด้านขวาของหลักเรียกว่า Leaf ซึ่งแผนภาพนี้สามารถทำได้เองโดยไม่ต้องใช้คอมพิวเตอร์

2.3.4 Normality Probability Plot เป็นเทคนิคที่ใช้กราฟทดสอบการแจกแจงข้อมูล ว่าข้อมูลมีการแจกแจงแบบปกติหรือไม่ โดยเปรียบเทียบข้อมูลที่ต้องการทดสอบเทียบกับการแจกแจงแบบปกติตามทฤษฎี

## 3. การทดสอบสมมติฐาน

สมมติฐาน คือ ข้อสมมติ หรือข้อความที่เกี่ยวข้องกับประชากรชุดเดียวหรือมากกว่า ซึ่งอาจ

เป็นจริงหรือไม่ก็ได้ โดยการประเมินผลทางวิธีทดสอบจะต้องตั้งสมมติฐานและทดสอบสมมติฐานหรือทดสอบนัยสำคัญ

- ถ้าการทดสอบมีนัยสำคัญ จะปฏิเสธสมมติฐานหลัก
- ถ้าการทดสอบไม่มีนัยสำคัญ จะยอมรับสมมติฐานหลัก

บริเวณวิกฤติ หมายถึงบริเวณที่จะทำให้เกิดการปฏิเสธสมมติฐาน โดยหาได้จากการเปิดตารางค่าสถิติ ซึ่งขึ้นอยู่กับที่ตั้งสมมติฐานและการเลือกระดับนัยสำคัญ

3.1 ขั้นตอนการทดสอบสมมติฐาน มี 6 ขั้นตอน ดังนี้

3.1.1 ตั้งสมมติฐาน โดยกำหนด  $H_0$  เป็นสมมติฐานหลัก และ  $H_1$  เป็นสมมติฐานรอง

3.1.2 เลือกระดับนัยสำคัญ ( $\alpha = 0.01$  หรือระดับความเชื่อมั่นที่ 99% จะเท่ากับ 0.99 หรือ  $\alpha = 0.05$  หรือระดับความเชื่อมั่นที่ 95% จะเท่ากับ 0.95) ซึ่งระดับนัยสำคัญจะหมายถึงความน่าจะเป็นในการปฏิเสธ  $H_0$  ซึ่ง  $H_0$  เป็นจริง เขียนแทนด้วยสัญลักษณ์  $\alpha$  และระดับความเชื่อมั่น หมายถึง ความน่าจะเป็นในการยอมรับ  $H_0$  ที่เป็นจริง เขียนแทนด้วย  $1 - \alpha$

3.1.3 เลือกสถิติที่จะทดสอบ เช่นทดสอบค่าเฉลี่ย ใช้สถิติทดสอบคือ t, ทดสอบค่าความแปรปรวน ใช้สถิติทดสอบคือ F เป็นต้น

3.1.4 กำหนดบริเวณวิกฤติ โดยเปิดจากตารางสถิติ

3.1.5 คำนวณค่าทางสถิติ จากข้อมูลการวัดตัวอย่างที่สุ่มมา

3.1.6 สรุปผล

- ถ้าค่าทางสถิติที่คำนวณได้อยู่ในบริเวณวิกฤติ เรียกว่า การทดสอบมีนัยสำคัญ จะปฏิเสธ  $H_0$  ยอมรับ  $H_1$

- ถ้าค่าทางสถิติที่คำนวณได้ไม่อยู่ในบริเวณวิกฤติ เรียกว่า การทดสอบไม่มีนัยสำคัญ จะยอมรับ  $H_0$  ปฏิเสธ  $H_1$

3.2 การทดสอบ 2 ทาง (Two - tail) เป็นการทดสอบโดยตั้งสมมติฐานที่ว่า  $H_0$  คือค่าทั้งสองไม่แตกต่างกัน และ  $H_1$  คือค่าทั้งสองแตกต่างกัน แทนด้วย  $H_0: \mu_1 = \mu_2$  และ  $H_1: \mu_1 \neq \mu_2$

3.3 การทดสอบทางเดียว (One - tail) เป็นการทดสอบโดยตั้งสมมติฐานที่ว่า  $H_0$  คือค่าทั้งสองไม่แตกต่างกัน และ  $H_1$  คือค่าของวิธีที่หนึ่งมากกว่าหรือน้อยกว่าวิธีที่สอง แทนด้วย

$H_0: \mu_1 = \mu_2$  และ  $H_1: \mu_1 > \mu_2$  หรือ  $H_1: \mu_1 < \mu_2$

#### 4. การทดสอบ Outlier

Outlier หมายถึง ข้อมูลที่มีค่าแตกต่างกันทั้งมากกว่าและน้อยกว่าจากข้อมูลในชุดเดียวกัน จนกระทั่งสงสัยว่าเป็นข้อมูลที่ไม่อยู่ในกลุ่มเดียวกัน ซึ่งเป็นสาเหตุทำให้ผลการวัดเช่นค่าเฉลี่ยหรือค่าการกระจายของข้อมูลคลาดเคลื่อน อาจมีสาเหตุมาจากการสูญเสียระหว่างเตรียมตัวอย่าง การปนเปื้อน เครื่องมือ/อุปกรณ์วัดมีปัญหา หรืออาจเกิดจากตัวผู้วิเคราะห์เองก็ได้ และเมื่อหาสาเหตุได้แล้ว ต้องดำเนินการแก้ไข บันทึกการแก้ไขและหาทางป้องกัน

การทดสอบ Outlier โดยวิธีของ Grubbs' test โดยการหาค่าอัตราส่วนค่าความแตกต่างระหว่างค่าที่สงสัยกับค่าเฉลี่ยของตัวอย่างกับค่าส่วนเบี่ยงเบนมาตรฐานที่ยังไม่ตัดค่าสงสัยออก และนำค่าที่ได้ไปเทียบกับค่าวิกฤติในตารางของ Grubbs' test ซึ่งสามารถคำนวณค่า  $G_{exp}$  ได้จากสูตร

$$G_{exp} = \frac{\bar{x} - x_i}{s} \quad \text{ใช้กรณีที่ค่าที่สงสัยอยู่ด้านต่ำสุด} \quad \text{หรือ}$$

$$G_{exp} = \frac{x_n - \bar{x}}{s} \quad \text{ใช้กรณีที่ค่าที่สงสัยอยู่ด้านสูงสุด}$$

ถ้า  $G_{exp} > G_{crit}$  สรุปได้ว่าค่าที่สงสัยเป็น Outlier เมื่อ  $G_{exp}$  เป็นค่าที่คำนวณได้ และ  $G_{crit}$  เป็นค่าที่เปิดจากตาราง

#### 5. การทดสอบแบบ F-test

เป็นการทดสอบเปรียบเทียบการกระจายของข้อมูล 2 ชุด เช่นการเปรียบเทียบความเที่ยงจากการวัดของข้อมูลที่แตกต่างกัน 2 ชุดข้อมูล ซึ่งขั้นตอนทดสอบแบบ F-test มีดังนี้

5.1 กำหนดให้  $H_0: \sigma_1 = \sigma_2$  และ กำหนดให้  $H_1: \sigma_1 \neq \sigma_2$

5.2 คำนวณหาค่าความแปรปรวนของข้อมูลทั้ง 2 ชุด

5.3 คำนวณค่าทางสถิติจากสูตร  $F = \frac{s_2^2}{s_1^2}$  เมื่อ Degree of freedom ( $V$ ) = n-1

5.4 นำค่าที่คำนวณได้เทียบกับค่าในตารางสถิติ

5.5 ถ้า  $F_{คำนวณ} < F_{วิกฤติ}$  ให้สรุปว่ายอมรับสมมติฐานหลัก คือชุดข้อมูลทั้ง 2 ชุดมีความเที่ยงที่ไม่แตกต่างกัน เป็นต้น

#### 6. การวิเคราะห์ความแปรปรวน (Analysis of Variance)

เป็นวิธีการที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร เมื่อจำนวนประชากรที่ต้องการทดสอบมีมากกว่า 2 หมู่ หรือใช้วิเคราะห์หาสาเหตุของการแปรผัน ซึ่งอาจมีปัจจัยเดียวหรือ

หลายปัจจัยก็ได้

การวิเคราะห์ความแปรปรวน จะพิจารณาจาก อัตราส่วนของความแปรปรวนระหว่างกลุ่มและความแปรปรวนภายในกลุ่มเดียวกัน พิจารณาความแตกต่างได้ดังนี้

- เมื่ออัตราส่วนมีค่าน้อย แสดงว่าความแปรปรวนระหว่างกลุ่มและความแปรปรวนภายในกลุ่มเดียวกัน มีค่าพอๆ กัน การทดสอบสมมติฐานจึงไม่นับสำคัญ ดังนั้นค่าเฉลี่ยของกลุ่มทุกกลุ่มไม่แตกต่างกัน

- เมื่ออัตราส่วนมีค่ามาก แสดงว่าความแปรปรวนระหว่างกลุ่มมีมาก เมื่อเทียบกับความแปรปรวนภายในกลุ่มเดียวกัน มีค่าพอๆ กัน การทดสอบสมมติฐานจึงมีนัยสำคัญ ดังนั้นค่าเฉลี่ยของประชากรต่างกัน โดยมีค่าเฉลี่ยอย่างน้อย 1 กลุ่มที่แตกต่างจากค่าเฉลี่ยอื่น

อนึ่ง การวิเคราะห์ความแปรปรวนนี้ สามารถใช้สูตรสำเร็จรูปจากโปรแกรม Excel ในการคำนวณได้ โดยเข้าไปที่ Tool/Data Analysis และเลือกเครื่องมือที่จะใช้คำนวณทางสถิติได้เลย เช่น F-test, t-test, ANOVAs แต่หากไม่มีโปรแกรม Data Analysis สามารถเพิ่มเข้าไปได้ โดยไปที่ Tool/Add-Ins/เลือกที่ Analysis Toolpak และ Analysis Toolpak-VBA/กด OK เครื่องจะทำการติดตั้งโปรแกรม Data Analysis ให้

จากความรู้ที่ได้กล่าวมานั้น สามารถนำความรู้ที่ได้มาใช้อ้างอิงเพื่อให้เกิดความน่าเชื่อถือทางด้านการควบคุมคุณภาพเพราะสามารถลดความผิดพลาดของกระบวนการทำงาน โดยลดปัจจัยที่ทำให้เกิดความแปรผัน ทำให้ผลที่ได้มีความน่าเชื่อถือ ถูกต้อง อีกทั้งยังสามารถใช้เป็นข้อมูลในการตัดสินใจในการบริหารคน การเลือกเครื่องมือ เลือกวิธีวิเคราะห์ และการแปรผลของข้อมูล เพราะฉะนั้นความรู้ทางด้านสถิติจึงเป็นสิ่งสำคัญเพื่อใช้สถิติในกระบวนการควบคุมคุณภาพสำหรับใช้ในการแปรผลการวิเคราะห์และการนำเสนอผลการวิเคราะห์ได้อย่างมีประสิทธิภาพ.

#### เอกสารอ้างอิง

เอกสารประกอบการฝึกอบรม – สัมมนาวิชาการ เรื่อง การคำนวณค่าทางสถิติสำหรับงานวิเคราะห์ทดสอบ ระหว่างวันที่ 13 – 14 มกราคม 2554 โดยสำนักพัฒนาศักยภาพนักวิทยาศาสตร์ห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ