

## คำจำกัดความ

### Alternative hypothesis

หนึ่งของสองข้อความที่ขัดแย้งกันเกี่ยวกับค่าของตัวพารามิเตอร์ที่สร้างขึ้นเพื่อทดสอบสมมติฐาน ข้อความอื่นเป็น null hypothesis alternative hypothesis ใช้แทนช่วงของค่าพารามิเตอร์ที่อาจเป็นไปได้ อาจเป็นข้างหนึ่ง ( $H_a : \mu < 15$ ) หรือสองข้าง ( $H_a : \mu \neq 15$ )

### รูประฆัง (bell shape)

รูปแบบข้อมูลที่พบบ่อย ๆ ชนิดหนึ่ง มีอยู่ค่าหนึ่งที่เกิดบ่อยที่สุด ส่วนเบี่ยงเบนในทิศทางใด ทิศทางหนึ่งจากศูนย์กลางลดลงเหมือนกัน การแจกแจงสมมาตร

### การทดลองเบอร์นูลลี (Bernoulli trial)

บันทึกผลลัพธ์ของการทดลองหนึ่งเป็น “สำเร็จ” หรือ “ไม่สำเร็จ” อย่างใดอย่างหนึ่ง ความน่าจะเป็นของความสำเร็จแสดงได้ด้วยพารามิเตอร์  $p$  มีหลาย ๆ การทดลองของชีวิตจริง สามารถประกอบด้วยอนุกรมของการทดลอง Bernoulli ผลลัพธ์ที่วัดได้ในการทดลองนี้เป็น  $X$  (จำนวนของความสำเร็จทั้งหมด) ถ้าหากว่าการทดลองมีความอิสระซึ่งกันและกัน และถ้าค่าของ  $p$  ยังคงที่ พฤติการณ์ที่น่าจะเป็นของ  $X$  อธิบายได้โดยการแจกแจงทวินาม

### ทฤษฎีขีดจำกัดส่วนกลาง (Central limit Theorem)

ทฤษฎีนี้กล่าวว่าพฤติกรรมของมัชฌิมเลขคณิตตัวอย่างประมาณได้ด้วยโค้งปกติ (ค่าสังเกตของตัวอย่างสุ่มได้มาจากประชากรที่มีความแปรปรวนที่จำกัด) ปรากฏจากทฤษฎีขีดจำกัดส่วนกลาง หลักของการอนุมานจะยากแก่การปฏิบัติ

### ช่วงความเชื่อมั่น (Confidence interval)

ช่วงของค่าที่มีความน่าจะเป็น (โดยทั่ว ๆ ไป .90 หรือ .95) ของการบรรจุพารามิเตอร์ที่ไม่ทราบ ช่วงความเชื่อมั่นใช้ในการอนุมานสถานะที่สภาพของค่าสำหรับพารามิเตอร์ในคำถามที่จะหยิบมาใช้ไม่ได้ ถ้าหากว่าค่านั้นหยิบมาใช้ได้ การวิเคราะห์จำเป็นต้องทดสอบสมมติฐานมากกว่า ช่วงความเชื่อมั่น

### สหสัมพันธ์ (Correlation)

แนวโน้มสำหรับสองตัวแปรที่มีการแจกแจงร่วมกันตามความสัมพันธ์ร่วมกัน ถ้าหากว่าค่า

มากของ X มีแนวโน้มสัมพันธ์กับค่ามากของ Y สหสัมพันธ์มีค่าเป็นบวก ถ้าหากว่าค่ามากของ X มีแนวโน้มสัมพันธ์กับค่าน้อยของ Y สหสัมพันธ์มีค่าเป็นลบ ขนาดของความสัมพันธ์เชิงเส้นวัดได้ด้วยสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง r ค่าของ r เข้าใกล้บวกหรือลบ 1 แสดงถึงสหสัมพันธ์บวกหรือลบมากตามลำดับ ค่าของ r เข้าใกล้ 0 แสดงถึงสหสัมพันธ์น้อย สำหรับกลุ่มใด ๆ ของจุดค่า r จะอยู่ระหว่าง  $-1 \leq r \leq +1$

### ค่าวิกฤต (Critical value)

จุดตามแกนสถิติที่ใช้ทดสอบแยกเขตยอมรับจากเขตปฏิเสธถ้าหากว่าสถิติที่ใช้ทดสอบอยู่ที่ปลายมากกว่าค่าวิกฤตก็จะปฏิเสธ null hypothesis การวางค่าวิกฤตกำหนดได้ด้วยลักษณะของ alternative hypothesis กับขนาดของระดับนัยสำคัญ

### หลายเหลี่ยมความถี่สะสม (Cumulative frequency polygon)

พลอตกราฟเกี่ยวกับความถี่สะสม (ตามแกนตั้ง) กับขอบเขตชั้นบน (ตามแกนนอน) โยงจุดด้วยเส้นตรง

### องศาแห่งความอิสระ (degrees of freedom)

ตัวพารามิเตอร์ของการแจกแจง t, F และ  $\chi^2$  สำหรับการประยุกต์ของการแจกแจง t และ F จำนวนขององศาแห่งความอิสระมีความสัมพันธ์กับขนาดตัวอย่าง ในการประยุกต์ของการแจกแจง  $\chi^2$  จำนวนขององศาแห่งความอิสระขึ้นอยู่กับจำนวนของแถวกับคอลัมน์ในตาราง Contingency แต่อย่างไรก็ตามสภาวะการใช้องศาแห่งความอิสระเพื่อแสดงว่าการแจกแจงเฉพาะชนิดไหน t หรือ F หรือ  $\chi^2$  ดีที่สุดที่อธิบายพฤติกรรมของสถิติที่ใช้ทดสอบ

### การกระจาย (dispersion)

คุณลักษณะแสดงออกพื้นฐานชนิดหนึ่งของกลุ่มใดกลุ่มหนึ่งของข้อมูลที่เป็นตัวเลขการกระจายแสดงถึงขนาดของการกระจายระหว่างค่าสังเกต นั่นคือจำนวนค่าสังเกตไม่เหมือนกันทั้งหมดโดยทั่ว ๆ ไปวัดออกมาในรูปของส่วนเบี่ยงเบนมาตรฐาน

### การแจกแจงความถี่ (frequency distribution)

ตารางแสดงจำนวนค่าสังเกตของตัวอย่างแต่ละชั้นของกลุ่มเฉพาะของหลาย ๆ ชั้น การสร้างการแจกแจงความถี่เป็นขั้นแรกของการยกกลุ่มข้อมูลดิบ ถ้าหากต้องการรูปภาพเพิ่มเติม ข่าวสารเดียวกันสามารถเสนอเป็นแผนภูมิหรือหลายเหลี่ยมความถี่

## หลายเหลี่ยมความถี่ (frequency polygon)

กราฟเกี่ยวกับความถี่ของชั้นหนึ่งหากแทนได้ด้วยจุดหนึ่งที่พลอตบนจุดกลางของชั้นนั้นแล้วโยงจุดด้วยเส้นตรง

## มัชฌิมเลขคณิตที่คำนวณได้จากแจกแจงความถี่เป็นกลุ่ม (grouped mean)

มัชฌิมเลขคณิตคำนวณได้จากกลุ่มข้อมูลที่จัดเป็นกลุ่มของชั้น สมมติให้แต่ละค่าสังเกตในชั้นที่กำหนดให้มีค่าเท่ากับจุดกลางของชั้น grouped standard deviation สามารถคำนวณได้ด้วยวิธีการเดียวกัน

## แผนภูมิ (histogram)

กราฟในที่ซึ่งใช้แทนความถี่ของชั้นหนึ่งด้วยความสูงของแท่งหนึ่ง

## ช่วงกึ่งควอไทล์ (interquartile range)

การวัดการกระจายกำหนดได้ด้วยผลต่างระหว่างเปอร์เซ็นต์ที่ 75 กับเปอร์เซ็นต์ที่ 25 ค่ามากของช่วงกึ่งควอไทล์แสดงว่าข้อมูลมีการเปลี่ยนแปลงสูง

## เกณฑ์กำลังสองน้อยที่สุด (least squares criterion)

เกณฑ์ที่ใช้กันอย่างกว้างขวางสำหรับการตัดสินใจว่าเส้นตรงเส้นไหนเหมาะที่สุดสำหรับกลุ่มข้อมูลที่กำหนดให้ ตามเกณฑ์เส้นที่เหมาะสมที่สุด  $y = a + bx$  ตลอด  $n$  จุด  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  เป็นเกณฑ์หนึ่งที่ทำให้ผลรวมของส่วนเบี่ยงเบนในแนวตั้งกำลังสองของจุดจากเส้นมีค่าน้อยที่สุดนั่นคือ

$$\sum_{i=1}^n (y_i - a - bx_i)^2$$

สูตรสำหรับค่าน้อยที่สุดของ  $a$  และ  $b$  กำหนดในทฤษฎีกำลังสองน้อยที่สุด เส้นที่สอดคล้องเกณฑ์กำลังสองน้อยที่สุดเรียกว่าเส้นถดถอยกำลังสองน้อยที่สุด

## ระดับนัยสำคัญ (level of significance)

ความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ในสภาวะการทดสอบสมมติฐานยอมรับสมมติฐานเว้นแต่ข้อมูลแสดงว่าผิดเกินข้อสงสัยที่สมเหตุสมผลทั้งหมด ระดับนัยสำคัญให้ข้อกำหนดเป็นตัวเลขของอะไรบางอย่างที่ประกอบข้อสงสัยที่สมเหตุสมผล ในตัวอย่างส่วนมากผู้ทดลองเป็นคนปรับให้เป็น .01 หรือ .05

## ความสัมพันธ์เชิงเส้น (linear relationship)

ความสัมพันธ์ระหว่างสองตัวแปรที่สามารถประมาณด้วยเส้นตรงได้เพียงพอ

### ที่ตั้ง (location)

คุณลักษณะแสดงออกพื้นฐานชนิดหนึ่งของความผันแปรทางสถิติที่ตั้งอ้างอิงถึงแนวโน้มเข้าสู่ส่วนกลางนั้นคือ ตำแหน่งตามสเกลที่วัดได้ที่ค่าสังเกตมีแนวทางตั้งอยู่ โดยทั่ว ๆ ไปแสดงออกในรูปของมัธยฐานเลขคณิตของตัวอย่าง

### ตัวแบบทางคณิตศาสตร์ (mathematical model)

ฟังก์ชันน่าจะเป็นที่ใช้ประมาณพฤติกรรมของปรากฏการณ์ที่กำหนดให้ ตัวแบบคณิตศาสตร์ที่สำคัญที่สุดคือเส้นโค้งปกติ

### null hypothesis

สำหรับปัญหาหนึ่งตัวอย่าง ข้อความที่ตัวพารามิเตอร์เท่ากับค่าที่กำหนด ( $H_0: \mu = \mu_0$ ) ถ้าหากว่าสองตัวอย่างหรือมากกว่า (ตัวพารามิเตอร์หลายตัว) null hypothesis มีข้อความว่าตัวพารามิเตอร์ (หลายตัว) เท่ากัน ( $H_0: p_1 = p_2$ )

### พารามิเตอร์ (parameter)

ตัวคงที่ที่ปรากฏในฟังก์ชันน่าจะเป็นที่กำหนดกลุ่มของการแจกแจง สองการแจกแจงที่อยู่ในกลุ่มที่มีค่าพารามิเตอร์ต่างกันจะต่างกันในรูปแบบที่ตั้งหรือการกระจายในการแจกแจงแบบ  $t$ ,  $F$  และ  $\chi^2$  มีเพียงพารามิเตอร์ตัวเดียวเรียกว่า องศาแห่งความอิสระ ในการแจกแจงปกติมีสองตัวพารามิเตอร์  $\mu$  และ  $\sigma$  การแจกแจงทวินามมีสองพารามิเตอร์ด้วย  $n$  และ  $p$

### เปอร์เซ็นต์ไทล์ (percentile)

ค่าตามสเกลที่วัดได้มีค่ามากกว่าเปอร์เซ็นต์ที่แน่นอนของค่าสังเกตทั้งหมดในตัวอย่าง (หรือประชากร) โดยทั่ว ๆ ไปเปอร์เซ็นต์ไทล์ที่  $p$  มีค่าเท่ากับ  $\alpha_p$  เป็นตัวเลขที่ซึ่ง  $p(X < \alpha_p) = p/100$  เปอร์เซ็นต์ไทล์ที่ 25, 50 และ 75 ในบางครั้งกล่าวออกมาในรูปควอไทล์ เปอร์เซ็นต์ที่ 50 เป็นที่ทราบในรูปของมัธยฐาน

### ส่วนเบี่ยงเบนมาตรฐานร่วม (pooled standard deviation)

ตัวค่าประมาณหนึ่งของการกระจายที่ขึ้นอยู่กับข่าวสารที่บรรจุในสองตัวอย่างสุ่มหรือมากกว่าที่มีความอิสระกัน ในปัญหาสองตัวอย่างสุ่ม เงื่อนไขทั่ว ๆ ไปคือ  $\sigma_1 = \sigma_2 (= \sigma)$  ส่วนเบี่ยงเบนมาตรฐานร่วม คือ ตัวค่าประมาณของ  $\sigma$

## การแจกแจงประชากร (population distribution)

กลุ่มของค่าที่วัดได้ทั้งหมดที่เป็นไปได้ของของชนิดหนึ่ง การแจกแจงประชากรเป็นอะไรชนิดหนึ่งที่ใช้ข้อมูลตัวอย่างมาแทน มีพารามิเตอร์และส่วนเบี่ยงเบนมาตรฐานของการแจกแจงประชากรแสดงได้โดย  $\mu$  และ  $\sigma$  ตามลำดับ

## ตัวแปรเชิงสุ่ม (random variable)

ฟังก์ชันหนึ่งของ  $X$  ที่สัมพันธ์กับจำนวนของแต่ละผลลัพธ์ของการทดลอง ถ้าหากว่าการทดลองทำซ้ำ ๆ กันหลาย ๆ ครั้ง ผลลัพธ์ควรจะแปรและค่าของตัวแปรเชิงสุ่มควรจะแปรด้วย เป็นที่ทราบว่าการแจกแจงของผลที่ได้เป็นเสมือนการแจกแจงความน่าจะเป็นของตัวแปรเชิงสุ่ม  $X$

## การแจกแจงตัวอย่าง (Sample distribution)

กลุ่มของค่าที่วัดได้เกี่ยวกับเรื่องของตัวอย่าง มีพารามิเตอร์และส่วนเบี่ยงเบนมาตรฐานของการแจกแจงตัวอย่างแสดงได้เป็น  $\bar{x}$  กับ  $s$  ตามลำดับ วัตถุประสงค์ของทุก ๆ กระบวนการอนุมานต้องเกิดจากอะไรที่เป็นจริงในการแจกแจงตัวอย่าง ต่ออะไรที่อาจเป็นจริงในการแจกแจงประชากร

## ตัวสถิติ (Statistic)

ปริมาณใดปริมาณหนึ่งที่ได้คำนวณจากข้อมูลของตัวอย่าง ในการทดสอบสมมติฐานรูปแบบเฉพาะของตัวสถิติที่ปรากฏในกฎการตัดสินใจก็คือตัวสถิติที่ใช้ทดสอบ

## ความคลาดเคลื่อนประเภทที่ 1 (Type I error)

การปฏิเสธ null hypothesis เมื่อ null hypothesis เป็นจริง ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ I คือระดับนัยสำคัญ  $\alpha$

## ความคลาดเคลื่อนประเภทที่สอง (Type II error)

การยอมรับ null hypothesis เมื่อ alternative hypothesis เป็นจริง ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ II แสดงได้ด้วย  $\beta$  สำหรับทุก ๆ ค่าของพารามิเตอร์รวมอยู่ภายใต้ alternative hypothesis  $\beta$  มีค่าต่าง ๆ กัน กราฟของ  $\beta$  เทียบกับค่าที่สมมติขึ้นของพารามิเตอร์เรียกว่า Operating characteristic Curve