30 formulars & menus of MS EXCEL for Statistical Analysis

การวิเคราะห์ Data เบื้องต้น ด้วย MS EXCEL

30 สูตรและชุดคำสั่ง

(ที่อาจจะยังไม่มีใครบอกคุณ)

- สำหรับผู้ใช้ MS Excel ระดับ Beginner ถึงผู้วิเคราะห์ Data ระดับ
  - Diagnostic Analytics
  - Predictive Analytics

ยุวดี เปรมวิชัย ISBN 987-616-630-397-1

279.-

การวิเคราะห์ Data เบื้องต้น ด้วย MS EXCEL

30 สูตรและชุดคำสั่ง

(ที่อาจจะยังไม่มีใครบอกคุณ)

การวิเคราะห์ Data เบื้องต้นด้วย MS Excel

โดย ยุวดี เปรมวิชัย

ชื่อเรื่อง การวิเคราะห์ Data เบื้องต้นด้วย MS Excel
(Fundamental Data Analytics with MS Excel)
ปีที่จัดทำ มกราคม 2569
ชื่อผู้แต่ง ยุวดี เปรมวิชัย **e-mail** : yuwatisit@gmail.com
ราคา e-book 279.- บาท
ISBN (e-book) 978-616-630-397-1
จำนวนหน้า 151 หน้า
พิสูจน์อักษร/ออกแบบปก/จัดทำโดย ยุวดี เปรมวิชัย

สารบัญ

สารบัญ (ต่อ)

สารบัญ (ต่อ)

บทนำ

ความหมายของการวิเคราะห์ Data

Data หรือ ข้อมูล หมายถึง ข้อเท็จจริงของกิจการหรือองค์กรที่ยังไม่ผ่านกระบวนการจัดการใดๆทั้งสิ้น Data มีได้หลายลักษณะอาจเป็น ตัวเลข รูปภาพ ตัวอักษร เสียง ฯลฯ ลักษณะของ Data ถูกนำมาจัดเป็นประเภทตามระดับความสำคัญเชิงตัวเลขของ Data เรียกว่า มาตรวัดของ Data (Data Measurement) หมายถึง ลักษณะของ Data ที่แบ่งตามระดับของความมีค่าเชิงจำนวนเลข ซึ่งมีความสำคัญเพราะจะแสดงลักษณะและความสามารถในการคำนวณค่าของตัวแปร (Variable) ซึ่ง Data แบ่งตามมาตรวัดเป็น 4 ระดับ ดังนี้

1) Nominal Scale หมายถึง มาตรวัดระดับต่ำที่สุดเป็นเพียงการแบ่ง Data เป็นกลุ่มใหญ่ ๆ เท่านั้น โดยใช้ตัวเลขหรือสัญลักษณ์เป็นเครื่องแสดงกลุ่ม ไม่สามารถเปรียบเทียบ หรือ นำไปคำนวณได้ ถือว่าแต่ละกลุ่มมีความเสมอภาคหรือเท่าเทียมกัน และค่าที่กำหนดให้ก็ไม่มีความหมายเชิงจำนวนเลข เป็นการกำหนดเพื่อความสะดวกในการลงรหัส Data เท่านั้น ไม่มีความหมายที่จะ บวก ลบ คูณ หารกัน เช่น เพศชาย ใช้ตัวเลข 1 หญิง ใช้เลข 2 หรือ ระดับการศึกษา ปริญญาเอก ปริญญาโท ปริญญาตรี ใช้เลข 1, 2, 3 เป็นต้น

2) Ordinal Scale หมายถึง มาตรวัดระดับที่เป็นการแบ่ง Data เป็นกลุ่มใหญ่ เปรียบเทียบได้ว่ากลุ่มใดดีกว่ากัน มีความแตกต่างกัน แต่วัดความแตกต่างเป็นจำนวนเลขไม่ได้ เป็นค่าตัวเลขเชิงคุณภาพเท่านั้น เช่น กลุ่มคนไข้แบ่งเป็นกลุ่ม 1 และ 0 โดยกำหนด 1 = รักษาแล้วได้ผล และ 0 = รักษาแล้วไม่ได้ผล หรือ ความคิดเห็น

1= ไม่เห็นด้วย 2=เห็นด้วยเล็กน้อย 3= เห็นด้วยปานกลาง 4 = เห็นด้วยมาก 5 =เห็นด้วยมากที่สุด เป็นต้น

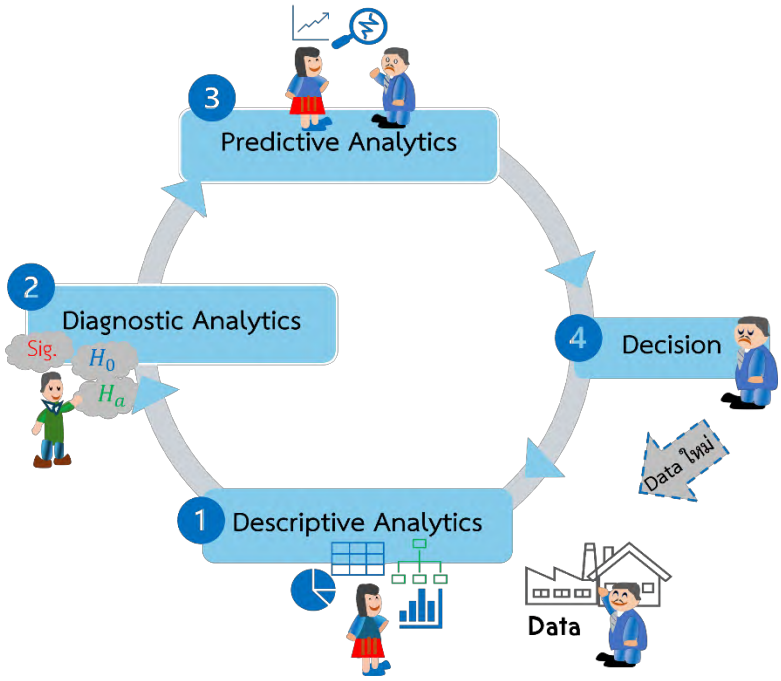
3) Interval Scale หมายถึง มาตรวัดระดับที่ค่าตัวเลขมีความหมายเพิ่มกว่ามาตรวัดระดับทั้งสองที่กล่าวมาแล้ว แต่ยังไม่ใช่ค่าจำนวนเลขที่สมบูรณ์ ไม่มีความหมายที่แท้จริง ได้แก่ อุณหภูมิ, เกรด, IQ เป็นต้น เช่น ณ อุณหภูมิ 0 องศา ไม่ได้หมายถึง ไม่มีอุณหภูมิ หรือ เกรดรายวิชา (Grade) มีค่า 0,1,2,3,4 เกรด 0 ไม่ได้หมายถึงไม่มีคะแนน เป็นต้น

4) Ratio Scale หมายถึง มาตรวัดระดับที่มีค่าตัวเลขเชิงปริมาณ เป็นมาตรวัดที่สมบูรณ์ที่สุด ค่าตัวเลขนั้นเป็นจำนวนเลขจริง เลข 0 มีค่าที่แท้จริงว่าไม่มีค่า ซึ่งการวิเคราะห์ทางสถิติส่วนใหญ่นิยมให้ Data เป็น Ratio Scale เพราะมีจุดเริ่มต้นที่แท้จริง และเปรียบเทียบด้วยการ บวก ลบ คูณ หาร กันได้ เช่น น้ำหนัก ความยาว อายุ รายได้ เป็นต้น

การวิเคราะห์ Data หรือ Data Analytics หมายถึง การใช้วิธีการทางสถิติหรือความรู้ทางสถิติวิเคราะห์ (Statistical Analysis) เพื่อค้นหาผลสรุปจาก Data ทั้งหลายที่มีอยู่ในกิจการใดกิจการหนึ่ง สำหรับให้เป็นประโยชน์ในการวางแผนหรือการตัดสินใจในกิจการนั้น โดยระดับของวิธีการวิเคราะห์ Data มี 4 ระดับ

1) Descriptive Analytics เป็นการวิเคราะห์ Data ขั้นพื้นฐาน เปรียบได้กับการที่พบ Data ครั้งแรก เป็นการวิเคราะห์เชิงบรรยายเพื่ออธิบายสถานการณ์ของกลุ่มข้อมูลนั้น ณ บริบทขณะนั้นเท่านั้น ไม่สามารถทำนายหรือตีความนอกเหนือไปถึงกลุ่มข้อมูลนั้นในบริบทอื่นๆ ได้ การวิเคราะห์ Data ระดับนี้ ได้แก่ การรวบรวม การตรวจสอบ และการนำเสนอ เป็นตาราง หรือ แผนภูมิ รวมถึงการวิเคราะห์ด้วยสถิติพรรณนา (Descriptive Statistics) ซึ่งเป็นสถิติขั้นพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) ร้อยละ(Percentage) ความถี่ (Frequency) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) มัธยฐาน (Median) เปอร์เซ็นต์ไทล์ (Percentile) เป็นต้น ซึ่งโปรแกรม MS Excel เป็นโปรแกรมที่ได้รับความนิยมในการใช้วิเคราะห์ Data ขั้นนี้อย่างมาก มีรูปแบบของ ตาราง กราฟ แผนภูมิ หลายประเภท และสามารถปรับแต่งได้ตามต้องการ

รวมถึงคุณสมบัติในการคำนวณของโปรแกรม MS Excel ด้วยสูตร (Formulas) ทำให้แสดงค่าสถิติพรรณนาได้สะดวก ใช้งานได้ง่าย



รูปที่ 1 ระดับของวิธีการวิเคราะห์ Data มี 4 ระดับ

2) Diagnostic Analytics เป็นการวิเคราะห์ Data เพื่อวิเคราะห์สาเหตุหรือข้อสงสัย หรือ ข้อสันนิษฐาน(Hypothesis) ที่มักจะต่อเนื่องมาจาก Descriptive Analytics วิธีการวิเคราะห์ Data ระดับนี้ คือ สถิติวิเคราะห์แบบสถิติอนุมาน (Inferential Statistics) เป็น การทดสอบสมมติฐาน (Hypothesis Testing)ที่องค์กรหรือกิจการใดๆมีความสงสัยต่อเนื่องมาจากผลของ Descriptive Analytics ซึ่งวิธีการของการทดสอบสมมติฐานมีค่าสถิติที่เกี่ยวข้องหลายค่า ตามลักษณะของสมมติฐานและลักษณะการกระจาย (Distribution)ของ Data นั้นๆ สถิติอนุมาน จำแนกได้ เป็น 2 ประเภท คือ

ประเภทที่ 1 สถิติแบบพาราเมตริกซ์ (Parametric Statistics) คือ วิธีการทางสถิติที่ให้ความสำคัญกับประชากรและการสุ่มตัวอย่าง ในการศึกษาสถิติอนุมานประเภทนี้กลุ่มตัวอย่าง (Sample) ที่นำมาศึกษาควรจะมีการแจกแจงปกติ ที่มีจำนวนมากพอ และตัวแปรมักจะมีมาตรวัด Interval Scale หรือ Ratio Scale

ประเภทที่ 2 สถิติแบบไม่ใช้พาราเมตริกซ์ (Nonparametric Statistics) คือ วิธีการทางสถิติที่ไม่เน้นลักษณะการแจกแจงปกติของประชากร และกลุ่มตัวอย่างมีขนาดเล็กไม่จำเป็นต้องมีจำนวนมาก ตัวแปรมีมาตรวัดระดับใดก็ได้ อาจเป็น Ordinal หรือ Nominal ก็ได้

โปรแกรม MS Excel มีเมนูคำสั่ง และสูตร(Formulars) สำหรับการวิเคราะห์ทางสถิติเบื้องต้นในขั้น Diagnostic Analytics นี้ เช่น หาค่าProbability ทดสอบ F-test ทดสอบ T-test เป็นต้น

3) Predictive Analytics เป็นการวิเคราะห์ Data เพื่อทำนายอนาคต หรือเรียกว่า การพยากรณ์ (Prediction) เกิดจากการศึกษาความสัมพันธ์ของ Data ที่เกิดขึ้นก่อนแล้วในช่วงเวลาที่ผ่านมา เพื่อนำไปใช้ทำนายเหตุการณ์ที่จะเกิดในอนาคต เช่น ศึกษาพบว่ายอดขายกาแฟของร้านกาแฟในห้างสรรพสินค้ามีความสัมพันธ์ทางบวกกับจำนวนลูกค้าของห้างสรรพสินค้านั้น ดังนั้นการตัดสินใจเปิดร้านกาแฟในห้างสรรพสินค้าที่มีลูกค้าจำนวนมาก จะทำนายยอดขายกาแฟล่วงหน้าได้ เป็นต้น วิธีการของสถิติวิเคราะห์เบื้องต้นในการศึกษาความสัมพันธ์ และทำนายข้อมูลในอนาคต คือ การวิเคราะห์สหสัมพันธ์และการถดถอย (Correlation and Regression Analysis) โปรแกรม MS Excel มีเมนูคำสั่ง และสูตร(Formulars) สามารถวิเคราะห์ Data ในระดับนี้ได้เช่นกัน

4) Prescriptive Analytics เป็นการวิเคราะห์ Data สำหรับการเตรียมการรับมือกับสิ่งที่คาดว่าจะเกิดขึ้นจากขั้นการทำนายของ Predictive Analytics ผลของการทำนายทำให้เกิดการตัดสินใจ (Decision) เตรียมการดำเนินกิจการต่างๆ เช่น ผลการทำนายยอดขายกาแฟปีหน้า พบว่า จะลดลง 50 % ดังนั้นการเตรียมการรับมือต่อยอดขายที่ลดลง จึงตัดสินใจให้มีการจัดโปรโมชั่นต่างๆ เป็นต้น ซึ่งจะพบว่า เมื่อถึงขั้น

ของ Prescriptive Analytics แล้ว เช่น มีการจัดโปรโมชั่น แล้ว ก็ทำให้เกิด Data ใหม่ขึ้นมาเพิ่มอีกจากกิจกรรมนั้นๆ ดังนั้นย่อมเกิดการวิเคราะห์ Data ขึ้นใหม่ เริ่ม Descriptive Analytics ของ Data ใหม่ นำไปสู่การทดสอบข้อสันนิษฐานอื่นๆ เกิดกระบวนการทดสอบสมมติฐาน และทำนาย ขึ้นอีกไม่สิ้นสุด แสดงให้เห็นว่าในทุกๆ องค์กรหรือทุกกิจการ มีกระบวนการของ Data Analytics ตลอดเวลา ไม่สามารถหยุดนิ่งได้ การวิเคราะห์ Data ในระดับนี้โปรแกรม MS Excel ทำหน้าที่ช่วยในการคำนวณต่าง ที่ต้องใช้ในการตัดสินใจได้อย่างดีตามความชำนาญด้านการคำนวณที่เป็นที่รู้จักกันดีของ MS Excel

คำศัพท์ที่สำคัญทางสถิติ

ในการใช้ MS Excel เพื่อวิเคราะห์สถิติเบื้องตัน หลายๆครั้งที่จำเป็นต้องกล่าวถึงคำศัพท์ทางสถิติ ดังนี้

1) ประชากร (Population) หมายถึงกลุ่มของ Data ทั้งหมด จะเป็นกลุ่มของ คน สัตว์ สิ่งของ ก็ได้ ประชากรในการวิเคราะห์ Data อาจมีกลุ่มเดียว หรือหลายกลุ่ม ก็ได้ และจำนวนของประชากรหรือเรียกว่าขนาดของประชากร นิยมใช้สัญลักษณ์เป็น N อาจหาจำนวนที่แน่นอนได้ หรือไม่สามารถหาจำนวนได้

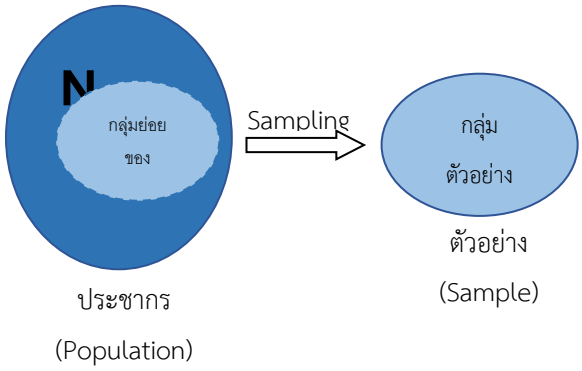
2) การแจกแจงความน่าจะเป็น (Probability Distribution) ในการศึกษาข้อมูลจำเป็นต้องศึกษาลักษณะของการแจกแจงความน่าจะเป็น (Probability Distribution) ของตัวแปรที่กำลังศึกษาก่อนเสมอ เพื่อเลือกใช้ประเภทของสถิติได้ถูกต้อง การแจกแจงความน่าจะเป็นแบ่งเป็น 2 ชนิด ตามลักษณะของตัวแปรที่ศึกษา คือ

(1) การแจกแจงความน่าจะเป็นชนิดไม่ต่อเนื่อง (Discrete Probability Distribution) หมายถึง การแจกแจงของตัวแปรที่ตัวแปรนั้นเป็นตัวเลขไม่ต่อเนื่องกัน เช่นตัวแปร X เป็นจำนวนคนไข้ที่รักษาโรคด้วยยาชนิดหนึ่งแล้วหายจากอาการโรคนี้ ดังนั้น X อาจจะมีค่า 0, 1, 2,... ก็ได้ เรียกว่า ตัวแปร X เป็นตัวแปรชนิดไม่ต่อเนื่อง ก็จะนำไปพิจารณาการแจกแจงของ X ได้หลายแบบตามธรรมชาติของ X นั้น ชื่อของ

การแจกแจงความน่าจะเป็นชนิดไม่ต่อเนื่อง เช่น การแจกแจงทวินาม (Binomial Distribution) การแจกแจงปัวซอง (Poisson Distribution) เป็นต้น

(2) การแจกแจงชนิดต่อเนื่อง (Continuous Probability Distribution) หมายถึง การแจกแจงของตัวแปรที่ตัวแปรนั้นเป็นตัวเลขที่ต่อเนื่องกัน เช่นตัวแปร X เป็นคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่มีคะแนนเต็ม 100 คะแนน ดังนั้น X อาจจะมีค่า 10, 11.5, 11.75, 12.0,... , 100 ก็ได้ เรียกว่า ตัวแปร X เป็นตัวแปรชนิดต่อเนื่อง ก็จะนำไปพิจารณาการแจกแจงของ X ได้หลายแบบตามธรรมชาติของ X นั้น ชื่อของการแจกแจงความน่าจะเป็นชนิดต่อเนื่อง เช่น การแจกแจงปกติ (Normal Distribution) การแจกแจง t (Student-T Distribution) เป็นต้น

3) กลุ่มตัวอย่าง หรือ ตัวอย่าง (Sample) หมายถึงบางส่วนหรือกลุ่มย่อยของประชากรทั้งหมด ถูกสุ่ม (Sampling) มาเป็น Data ในการศึกษาวิเคราะห์ ใช้สัญลักษณ์ขนาดของกลุ่มตัวอย่าง เป็น n



รูปที่ 2 ประชากร และ ตัวอย่าง

การที่ต้องใช้กลุ่มตัวอย่างมาศึกษาแทนประชากรทั้งหมดนั้น เพื่อเป็นการประหยัดทรัพยากร ได้แก่ ลดค่าใช้จ่าย ลดแรงงาน ลดเวลา เป็นต้น และบางกรณีมีความจำเป็นอย่างยิ่งที่หา Data ประชากรทั้งหมดไม่ได้จริงๆเช่น ศึกษาเม็ดเลือดขาวในกลุ่มคนไข้ชนิดหนึ่ง เป็นต้น นอกจากนี้ความเปลี่ยนแปลงอย่างรวดเร็วของข้อมูลในยุค

ปัจจุบัน ก็มีผลต่อการต้องรีบศึกษาในเวลาเร่งด่วน ทำให้ไม่สามารถศึกษาทั้งประชากรได้ การวิเคราะห์ Data ส่วนใหญ่ให้ความสำคัญที่การแจกแจงว่าข้อมูลที่สุ่มมาเป็นกลุ่มตัวอย่างควรมีการแจกแจงแบบ Normal Distribution มีหลักฐานที่เชื่อถือได้เกี่ยวกับการกำหนดขนาดตัวอย่าง (n) ที่เหมาะสม ศึกษาโดย Zikmund ในปี ค.ศ.1999 โดยยึดจาก ทฤษฎีแนวโน้มสู่ส่วนกลาง (Central Limit Theorem) ว่าขนาดของกลุ่มตัวอย่าง (n) ควรมีขนาดตั้งแต่ 30 หน่วย ($n \geq 30$) จึงจะเหมาะสมที่อนุโลมให้ว่าข้อมูลมีการแจกแจงแบบ Normal Distribution

4) ค่าพารามิเตอร์(Parameter) และ ค่าสถิติ(Statistic) ประชากรและกลุ่มตัวอย่าง จะมีการแสดงค่าสถิติต่าง ๆ ได้เช่นเดียวกันทุกค่า โดยค่าสถิติของประชากรเรียกว่า ค่าพารามิเตอร์ (Parameter) และค่าสถิติของกลุ่มตัวอย่างเรียกว่าค่าสถิติ(Statistic) นิยมใช้สัญลักษณ์ที่แตกต่างกันให้สังเกตได้ ดังตารางที่ 1

ตารางที่ 1 ค่าพารามิเตอร์(Parameter) และค่าสถิติ (Statistic)

ค่าพารามิเตอร์	ค่าสถิติ	ความหมาย
μ	$\bar{x}$	ค่าเฉลี่ย (Mean, Average)
σ	s หรือ sd	ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)
σ^2	s^2	ความแปรปรวน (Variance)
p	$\hat{p}$ (อ่านว่า p-cap)	ค่าสัดส่วน (Proportion)
ρ (อ่านว่า rho)	r	ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)
Y	$\hat{Y}$ (อ่านว่าY-cap)	ตัวแปร Response ของสมการถดถอย(Regression Equation)
β (อ่านว่า beta)	b	สัมประสิทธิ์ตัวแปรถดถอย (Coefficient of Regression Variable)

5) สมมติฐานทางสถิติ (Statistical Hypothesis) หมายถึงข้อสันนิษฐาน หรือ ความเชื่อของลักษณะประชากร ซึ่งอาจจะเป็นข้อสรุปที่เป็นจริงหรือไม่จริงก็ได้ เป็นข้อสมมติที่เชื่อหรือคาดว่าจะเกิดขึ้นในประชากร เมื่อแปลความด้วยข้อความทางสถิติ เรียกว่า สมมติฐานทางสถิติ เช่น นักการตลาดทำการศึกษายอดขายสินค้าชนิดหนึ่งที่มีวิธีโฆษณาแตกต่างกัน 2 วิธี คือ วิธี A และ วิธี B โดยสังเกตข้อมูลจากยอดขายเฉลี่ยในไตรมาศแรกของปีนี้ จากวิธี A ได้ยอดขายเฉลี่ยเป็น $\bar{X}_A$ และ จากวิธี B ได้ยอดขายเฉลี่ยเป็น $\bar{X}_B$ ตั้งสมมติฐานทางสถิติได้เป็นกรณีต่างๆ ดังนี้

$\mu_A = \mu_B$ หมายถึง ยอดขายเฉลี่ยในไตรมาศแรกที่ได้มาจากวิธีการโฆษณาทั้ง 2 วิธีเท่ากัน หรือไม่แตกต่างกัน

$\mu_A < \mu_B$ หมายถึง ยอดขายเฉลี่ยในไตรมาศแรกที่ได้มาจากการโฆษณาวิธี B มากกว่าวิธี A

$\mu_A > \mu_B$ หมายถึง ยอดขายเฉลี่ยในไตรมาศแรกได้มาจากการโฆษณาวิธี A มากกว่าวิธี B

$\mu_A \neq \mu_B$ หมายถึง ยอดขายเฉลี่ยในไตรมาศแรกที่ได้มาจากวิธีการโฆษณาทั้ง 2 วิธี แตกต่างกัน

สมมติฐานทางสถิติ ต้องกำหนดไว้ 2 สมมติฐาน คือ

(1) สมมติฐานหลัก (Null Hypothesis) สัญลักษณ์คือ H_0 เป็นข้อสมมติที่ใช้เป็นหลักเนื่องจากคาดว่า ข้อสมมตินี้มีค่าที่เกี่ยวข้องกับประชากรที่เป็นประเด็นของการวิจัยอย่างแท้จริง เป็นค่าที่สำคัญนำไปใช้คำนวณหาค่าเกณฑ์ตัดสินใจ เพื่อหาผลสรุปว่าจะปฏิเสธ (Reject) หรือไม่ปฏิเสธ (Not Reject) สมมติฐาน H_0 นั้น ดังนั้นการพิจารณาตั้ง H_0 จึงกำหนดเป็นข้อความหรือสมการที่มีคำหรือเครื่องหมายแสดงการเท่ากับ (=) ประกอบอยู่ด้วยเสมอ ทั้งนี้เพื่อความแน่นอนในการนำค่านี้ไปใช้ในการคำนวณต่อไป

(2) สมมติฐานรอง (Alternative Hypothesis) หรือสมมติฐานแย้ง สัญลักษณ์ คือ H_a เป็นข้อสมมติที่ตั้งไว้เพื่อแสดงความหมายตรงข้ามกับสมมติฐานหลัก ดังนี้

(2.1) $H_0 : \mu_A = \mu_B$

$H_a : \mu_A \neq \mu_B$

(2.2) $H_0 : \mu_A = \mu_B$

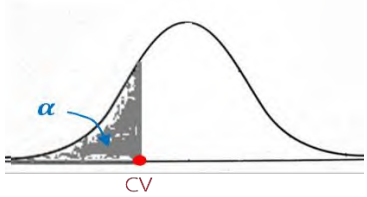
$H_a : \mu_A > \mu_B$

(2.3) $H_0 : \mu_A = \mu_B$

$H_a : \mu_A < \mu_B$

6) ระดับนัยสำคัญ (level of Significant) จะใช้สัญลักษณ์ α เป็นค่าแสดงให้เห็นว่าผลสรุปของการทดสอบสมมติฐาน เชื่อถือได้มากน้อยเพียงใด เช่น α = 0.05 หมายความว่า ผลสรุปของการทดสอบมีระดับนัยสำคัญ 0.05 มีความน่าเชื่อถือได้ในระดับความเชื่อมั่น (Confidence Level) เป็น 95% เป็นต้น หรือเรียกว่ามีสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient) เป็น 95%

7) ค่าวิกฤติ (Critical Value ; CV) หมายถึงค่าขอบเขตของระดับความเชื่อมั่นมีค่าตามระดับนัยสำคัญ α และตามชนิดของการแจกแจง



รูปที่ 3 ค่าวิกฤติ (Critical Value ; CV) และระดับนัยสำคัญ α

8) บริเวณวิกฤติ (Critical Region ; CR) หมายถึงความน่าจะเป็นหรือขนาดของพื้นที่ใต้เส้นโค้งของแต่ละ Distribution ตามสมมติฐานรอง และมีขนาดตามค่าของระดับนัยสำคัญ α ดังนี้

(1) $H_0 : \mu_A = \mu_B$

$H_a : \mu_A \neq \mu_B$