หน่วยที่ 9 การแจกแจงความน่าจะเป็น

เซต (set) หมายถึง กลุ่มของสิ่งต่าง ๆ ซึ่งมีคุณสมบัติที่แน่ชัด โดยที่สามารถบอกได้ว่าสิ่งใดสิ่งหนึ่งอยู่ในเซตที่กล่าวถึงหรือไม่ เรียกสิ่งที่อยู่ในเซตว่าสมาชิก (element หรือ member) เซตที่สามารถนับหรือบอกจำนวนสมาชิกได้ว่ามีเท่าไร คือนับสิ้นสุดได้เรียกว่า เซตจำกัด (finite set) เช่น เซตของเดือนในรอบปี เซตของโรงพยาบาลชุมชนในปี พ.ศ. 2547 ส่วนเซตที่ไม่สามารถบอกจำนวนสมาชิกได้เรียกว่า เซตอนันต์ (infinite set) เช่นเซตของจำนวนนับ เซตของจุดบนเส้นตรง
การเขียนเซตนิยมใช้อักษรภาษาอังกฤษตัวพิมพ์ใหญ่ ส่วน สมาชิกใช้ตัวคัวพิมพ์เล็ก

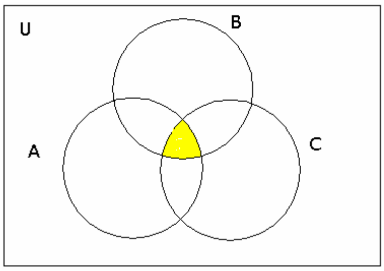
ชนิดของเซต

1. เอกภพสัมพัทธ์ (universal set หรือ universe) คือเซตที่รวมสมาชิกทั้งหมดที่อยู่ในขอบข่ายที่สนใจศึกษา เขียนแทนด้วย สัญลักษณ์ U เช่นศึกษาเรื่องการใช้บริการที่โรงพยาบาลชุมชนของชาวบ้าน หมู่บ้านห้วยเรียน เอกภพสัมพัทธ์ คือ ชาวบ้านทุกคนที่หมู่บ้านนี้

2. เซตว่าง (empty set หรือ null set) คือเซตที่ไม่มีสมาชิกใด ๆ อยู่เลย เขียนแทนด้วนสัญลักษณ์ { } หรือแทนด้วยสัญลักษณ์ $\varnothing$ อ่านว่าฟี (phi) เช่นเซตของอักษร “ ก ” ในคำว่าวิทยาศาสตร์สุขภาพ

3. เซตย่อย (subset) B เรียกว่าเป็นเซตย่อยของ A ก็ต่อเมื่อ ทุก ๆ สมาชิกของ A เป็นสมาชิกของ B เช่น เซตของเลขที่หมู่บ้านในตำบลเวียงเหนือคือ A = { 1,2,3,4,5,6 } เซตของหมู่บ้านที่เป็นเลขคู่คือ B = { 2,4 ,6 }

แผนภาพเวนน์ (Venn diagram) คือ แผนภาพเกี่ยวกับเซตที่แสดงให้เห็นถึงความสัมพันธ์หรือการดำเนินการของเซต ใช้รูปภาพปิดใด ๆ ก็ได้แทนเซต เช่น รูปวงกลม รูปสี่เหลี่ยม



แผนภาพเวนน์แสดง $A \cap B \cap C$ (พื้นที่แรเงา)

การดำเนินการของเซต (set operations)

1. ยูเนียน (union) ยูเนียนของเซต A และ B คือ เซตที่มีสมาชิกที่เป็นสมาชิกของเซต A หรือ สมาชิกของ เซต B หรือทั้งสองเซต เขียนแทนด้วยสัญลักษณ์ A U B (อ่านว่า A ยูเนียน B)

ตัวอย่าง ที่คลินิกจิตเวชมีคนไข้ทั้งหมด 12 คน คนไข้คนที่ 1, 2, 3, 4, 5 และ 6 ได้รับการรักษาโดยวิธีการให้ยารับประทาน ส่วนคนไข้คนที่ 2,4,7,8,9,10,11 และ 12 ได้รับการรักษาโดยวิธีการแนะนำ
กำหนดเซตดังนี้

ให้ A เป็นเซตคนไข้ที่ได้รับการรักษาโดยวิธีให้ยารับประทาน
ให้ B เป็นเซตคนไข้ที่ได้รับการรักษาโดยวิธีการแนะนำให้คำปรึกษา

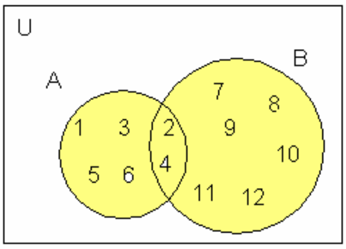
A = {คนไข้คนที่ 1, 2, 3, 4, 5, 6 }

B = { คนไข้คนที่ 2, 4, 7, 8, 9, 10, 11, 12 }

ดังนั้น A U B หมายถึงคนไข้ที่ได้รับการรักษาโดยวิธีการให้ยารับประทาน หรือ วิธีการแนะนำให้คำปรึกษา

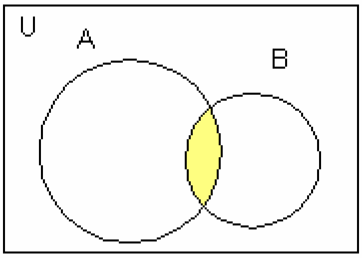
A U B = { คนไข้คนที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 }

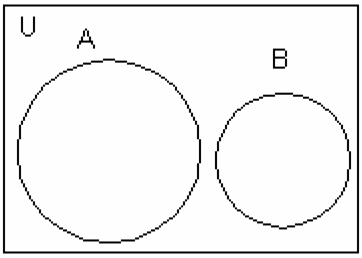
(ตัวอย่างหน้า 86)

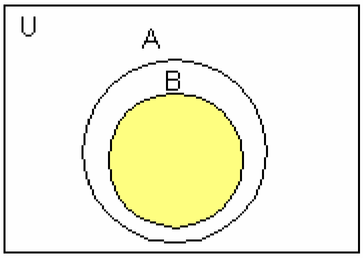


เขียนแผนภาพเวนน์แสดง A U B ได้ดังภาพ

2. อินเตอร์เซกชัน (intersection) อินเตอร์เซกชันของเซต A และ B คือ เซตที่มีสมาชิกที่เป็นสมาชิกของทั้งเซต A และ B แทนด้วยสัญลักษณ์ $A \cap B$ (อ่านว่า A อินเตอร์เซก B) โดยใช้สัญลักษณ์ $\cap$ แทนอินเตอร์เซกชัน







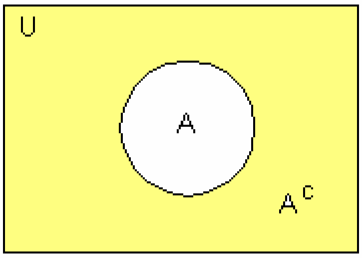
$A \cap B$ คือพื้นที่ตรงส่วนที่แรเงา

$A \cap B = \{\ \}$ หรือ $\varnothing$

ถ้า B เป็นเซตย่อยของ A จะได้ว่า $A \cap B = B$ (คือพื้นที่ใน B)

(ตัวอย่าง หน้า 88)

3. คอมพีเมนต์ (complement) คอมพีเมนต์ของเซต A ใด ๆ เมื่อเทียบกับเอกภพสัมพันธ์ U คือเซตของทุกสมาชิกที่อยู่ใน U แต่ไม่อยู่ใน A เขียนแทนด้วยสัญลักษณ์ A^C



แผนภาพเวนน์แสดง A^C (พื้นที่สีเหลือง)

ตัวอย่าง หญิงตั้งครรภ์ 50 คน ได้รับการดูแลก่อนคลอดที่โรงพยาบาลศูนย์ หญิงกลุ่มนี้มีอยู่ 4 คนมีเลือดกลุ่ม AB ดังนั้นที่เหลือ 46 คนมีเลือดกลุ่มอื่น ๆ ที่ไม่ใช่เลือดกลุ่ม AB

ให้ A = {x l x หญิงตั้งครรภ์ที่มีเลือดกลุ่ม AB }

ดังนั้น A^C หมายถึงหญิงตั้งครรภ์ที่มีเลือดกลุ่มอื่น ๆ ที่ไม่ใช่เลือดกลุ่ม AB ในหญิงตั้งครรภ์ 50 คนที่ได้รับการดูแลก่อนคลอดที่โรงพยาบาลศูนย์

A^C = { x l x หญิงตั้งครรภ์ที่มีเลือดกลุ่ม A , B และ O }

การเรียงสับเปลี่ยนและการจัดหมู่ เป็นเรื่องราวของจำนวนวิธีการจัดสิ่งของที่แตกต่างกัน โดยดูอันดับสิ่งของเรียกว่าการเรียงสับเปลี่ยน และเมื่อไม่ดูอันดับสิ่งของเรียกว่าการจัดหมู่ (หน้า 91)

การทดลองเชิงสุ่ม (random experiment) หมายถึง การทดลองที่ไม่สามารถพยากรณ์ผลลัพธ์ได้อย่างถูกต้องแน่นอน เนื่องจากผลลัพธ์อาจเกิดขึ้นได้หลายอย่าง เช่น การทอดลูกเต๋าเที่ยงตรงหนึ่งลูก ผลลัพธ์ที่สนใจคือจำนวนแต้มที่ปรากฏบนหน้าของลูกเต๋า ลูกเต๋าอาจขึ้นหน้า 1 , 2 , 3, 4, 5 หรือ 6 ก็ได้ ผลขอลการทอดลูกเต๋าจะปรากฏได้จากการทอดในแต่ละครั้ง ซึ่งผู้ทอดลูกเต๋าไม่สามารถทราบก่อนล่วงหน้าว่าลูกเต๋าจะขึ้นหน้าใด

ปริภูมิตัวอย่าง (sample space) หมายถึงเซตที่มีสมาชิกของผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองเชิงสุ่ม เขียนแทนด้วยสัญลักษณ์ S การทดลองเชิงสุ่มเดียวกันอาจมีปริภูมิตัวอย่างมากกว่าหนึ่งอย่างก็ได้ ทั้งนี้ขึ้นอยู่กับผลลัพธ์ที่สนใจ เช่น ผลลัพธ์

ที่สนใจคือจำนวนแต้มที่ปรากฏบนหน้าของลูกเต๋า ลูกเต๋าอาจขึ้นเป็นหน้า 1, 2, 3, 4, 5 หรือ 6 ก็ได้ นั่นคือ เซต S_1 ประกอบด้วยสมาชิกดังนี้ $S_1 = \{ 1, 2, 3, 4, 5, 6 \}$ ถ้าสนใจว่าผลที่ได้จากการทอดลูกเต๋าเป็นแต้มเลขคี่ หรือเลขคู่ จะได้ ว่าเซตของปริภูมิตัวอย่างซึ่งเขียนแทนด้วยเซต S_2 ประกอบด้วยสมาชิกดังนี้ S_2 = { แต้มเลขคี่ , แต้มเลขคู่ }

เหตุการณ์ (even) หมายถึงเซตย่อยของปริภูมิตัวอย่าง เหตุการณ์เขียนในรูปเซตจากผลลัพธ์ที่สนใจในการทดลองเชิงสุ่ม ตัวอย่างเช่น การทอดลูกเต๋าเที่ยงตรงหนึ่งลูก ผลลัพธ์ที่สนใจคือจำนวนแต้มที่ปรากฏบนหน้าของลูกเต๋าดังนั้น $S = \{1, 2, 3, 4, 5, 6 \}$ สนใจเหตุการณ์ E_1 , E_2, E_3 และ E_4 ดังต่อไปนี้

ถ้า E_1 เป็นเหตุการณ์ที่ได้แต้มซึ่งหารด้วย 2 ลงตัว จะได้ $E_1 = \{ 2, 4, 6 \}$

ถ้า E_2 เป็นเหตุการณ์ที่ได้แต้มต่ำกว่า 4 จะได้ $E_2 = \{ 1, 2, 3 \}$

ถ้า E_3 เป็นเหตุการณ์ที่ได้แต้มตั้งแต่ 3 ขึ้นไป จะได้ $E_3 = \{ 3, 4, 5, 6 \}$

ถ้า E_4 เป็นเหตุการณ์ที่ได้แต้มมากกว่า 6 จะได้ $E_4 = \{ \}$ หรือเขียนว่า $\varnothing$

1. ยูเนียนของเหตุการณ์ (union of events) ยูเนียนของเหตุการณ์ E_1 และ E_2 หมายถึง หมายถึง เหตุการณ์ซึ่งประกอบด้วยสมาชิกของเหตุการณ์ E_1 หรือ E_2 หรือทั้งสองเหตุการณ์ เขียนแทนด้วยสัญลักษณ์ E1 U E2

2. เหตุการณ์ร่วม (intersection of events) เหตุการณ์ร่วมของเหตุการณ์ E_1 และ E_2 หมายถึง เหตุการณ์ซึ่งประกอบด้วยสมาชิกที่อยู่ในเหตุการณ์ E_1 และ E_2 เขียนแทนด้วยสัญลักษณ์ $E_1 \cap E_2$ หรือเขียนว่า E_1E_2

3. เหตุการณ์ที่ไม่มีส่วนร่วม (mutually exclusive events หรือ disjoint events) เมื่อ E_1 และ E_2 เป็นเหตุการณ์สองเหตุการณ์ และ $E_1 \cap E_2 = \varnothing$ เรียกเหตุการณ์ E_1 และ E_2 ว่าเหตุการณ์ที่ไม่มีส่วนร่วม

4. คอมพลีเมนต์ของเหตุการณ์ (complement of event) เมื่อ E เป็นเหตุการณ์ที่มี S เป็นปริภูมิตัวอย่างของการทดลองเชิงสุ่ม คอมพลีเมนต์ของเหตุการณ์ E หมายถึง เหตุการณ์ที่ประกอบด้วยสมาชิกที่อยู่ในปริภูมิตัวอย่าง S แต่ไม่อยู่ในเหตุการณ์ E เขียนแทนด้วยสัญลักษณ์ E^C

ตัวอย่าง การทอดลูกเต๋าเที่ยงตรงหนึ่งลูก ผลลัพธ์ที่สนใจคือจำนวนแต้มที่ปรากฏบนหน้าของลูกเต๋าดังนั้นได้

$S = \{ 1, 2, 3, 4, 5, 6 \}$ สนใจเหตุการณ์ E_1 และ E_2 ดังนี้

E_1 เป็นเหตุการณ์ที่ได้แต้มซึ่งหารด้วย 3 ลงตัว จะได้ $E_1 = \{ 3, 6\}$

E_2 เป็นเหตุการณ์ที่ได้แต้มมากกว่า 4 จะได้ $E_2 = \{ 5, 6 \}$

ให้หาเหตุการณ์ต่อไปนี้

2.1 $E_1 \cup E_2 = \{3, 5, 6\}$

2.2 $E_1 \cap E_2 = \{6\}$

2.3 $E^C_1 = \{1, 2, 4, 5\}$

2.4 $E^C_2 = \{1, 2, 3, 4 \}$

ความน่าจะเป็น (probability) หมายถึง ค่าความถี่สัมพัทธ์ ที่บอกให้ทราบว่าเหตุการณ์ที่สนใจมีโอกาสเกิดขึ้นได้มากน้อยเพียงใด (ตัวอย่างคำนวณหน้า 98)

กิจกรรม 1. ความคิดเห็นของประชาชนในเขตบางพลัดเกี่ยวกับความคิดเห็นต่อการงดสูบบุหรี่ในสวนสาธารณะสุ่มตัวอย่างประชาชนในแต่ละแขวงของเขตนี้รวมจำนวน 500 คน ได้ข้อมูลดังตาราง

แขวง	ความคิดเห็น			รวม
	เห็นด้วยอย่างยิ่ง (A)	เห็นด้วย (B)	ไม่เห็นด้วย (C)	
บางบำหรุ (D)	100	20	5	125
บางพลัด (E)	115	5	5	125
บางอ้อ (F)	50	60	15	125
บางยี่ขัน (G)	35	50	40	125
รวม	300	135	65	500

ให้ A เป็นเหตุการณ์ที่ผู้ตอบเห็นด้วยอย่างยิ่งต่อการสูบบุหรี่

ให้ B เป็นเหตุการณ์ที่ผู้ตอบเห็นด้วยต่อการสูบบุหรี่

ให้ C เป็นเหตุการณ์ที่ผู้ตอบไม่เห็นด้วยต่อการสูบบุหรี่

ให้ D เป็นเหตุการณ์ที่ผู้ตอบอยู่เขตบางบำหรุ

ให้ E เป็นเหตุการณ์ที่ผู้ตอบอยู่เขตบางพลัด

ให้ F เป็นเหตุการณ์ที่ผู้ตอบอยู่เขตบางอ้อ

ให้ G เป็นเหตุการณ์ที่ผู้ตอบอยู่เขตบางยี่ขัน

ถ้าสุ่มตัวอย่างประชาชนมา 1 คน จาก 500 คนนี้ จงหาค่าความน่าจะเป็นในข้อต่อไปนี้

1.1 เห็นด้วยอย่างยิ่งต่อการงดสูบบุหรี่ P(A) = 300/500 = 0.60

1.2 ผู้ตอบอยู่แขวงบางบำหรุ P(D) = 125/500 = 0.25

1.3 ตอบว่าไม่เห็นด้วย หรือมีภูมิลำเนาอยู่แขวงบางยี่ขัน P(C U G) = ((65 + (125-40)) /500 = 0.30

1.4 เห็นด้วย หรือไม่เห็นด้วยต่อการงดสูบบุหรี่ P(B U C) = (135+65) / 500 = 0.40

1.5 อยู่ในเขตบางพลัด และเห็นด้วยต่อการงดสูบบุหรี่ $P(E \cap B) = 5/500 = 0.01$

1.6 อยู่ในเขตบางพลัด P(E) = 125/500 = 0.25

การแจกแจงความน่าจะเป็น หมายถึง การแจกแจงความถี่ของตัวแปรสุ่มชนิดไม่ต่อเนื่อง และชนิดต่อเนื่อง

ตัวแปรสุ่ม (random variable) คือ ฟังก์ชันซึ่งเปลี่ยนทุก ๆ เหตุการณ์ ในปริภูมิตัวอย่างให้เป็นตัวเลข การเขียนตัวแปรสุ่มใช้อักษรภาษาอังกฤษตัวพิมพ์ใหญ่แทนตัวแปร และค่าของตัวแปรใช้อักษรภาษาอังกฤษตัวพิมพ์เล็กตัวอักษรเดียวกัน

ชนิดตัวแปรสุ่ม

1. ตัวแปรสุ่มชนิดไม่ต่อเนื่อง หมายถึง ค่าของตัวแปรสุ่มที่ถูกกำหนดได้ด้วยจำนวนนับใด ๆ เช่น Y เป็นจำนวนคนที่เคยมารับบริการรักษาพยาบาลในโครงการสร้างหลักประกันสุขภาพถ้วนหน้าจากผู้มีบัตรประกันสุขภาพถ้วนหน้า 4 คน เพราะฉะนั้นค่าของตัวแปรคือ y = 0, 1, 2, 3, 4

2. ตัวแปรสุ่มชนิดต่อเนื่อง หมายถึง ค่าของตัวแปรสุ่มที่มีค่าเป็นตัวเลขใด ๆ ในช่วงที่ต่อเนื่องกัน และค่าของตัวแปรสุ่มในช่วงนั้น ๆ ไม่สามารถนับได้ว่ามีจำนวนเท่าใด คือเป็นจำนวนอนันต์ ค่าของตัวแปรเป็นค่าที่ได้จากการชั่ง ตวง วัด เช่น Z เป็นค่าความดันโลหิตระยะหัวใจบีบตัว (systolic blood pressure) ของคนกลุ่มอายุ 20-30 ปี ค่านี้ได้จากการวัด มีค่าอยู่ระหว่าง 121.8-128.2 มม. ปรอท เพราะฉะนั้นค่าของตัวแปรคือ z ซึ่ง $121.8 \leqq z \leqq 128.2$ มม. ปรอท

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม หมายถึง การแจกแจงความถี่ของตัวแปรสุ่มในค่าความน่าจะเป็น เป็นการดูว่าตัวแปรสุ่มค่าหนึ่ง ๆ มีค่าความน่าจะเป็นเท่ากับค่าใด ทำให้เห็นถึงการแจกแจงของตัวแปรสุ่มที่มีทั้งชนิดไม่ต่อเนื่องและต่อเนื่องนั้น ๆ ว่าแจกแจงในรูปลักษณะใด ด้วยฟังก์ชันความหนาแน่นของความน่าจะเป็นอย่างไร (หน้า 105)

ความหมายของการแจกแจงความน่าจะเป็น คือ การแจกแจงความถี่ของตัวแปรสุ่มในค่าความน่าจะเป็น ทำให้เห็นรูปลักษณะการแจกแจง

การแจกแจงทวินาม (binomial distribution) คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง ที่ผลการทดลองมี 2 อย่างคือ สำเร็จ และล้มเหลว เช่น การโยนเหรียญเที่ยงตรง 1 เหรียญ ให้เหรียญขึ้นหัวเป็นความสำเร็จ เหรียญขึ้นก้อยเป็นความล้มเหลว มาจากการทดลองที่เรียกว่าการทดลองแบร์นูลลี (ชื่อนักคณิตศาสตร์ชาวสวิท : เจมส์ แบร์นูลลี) (หน้า 109)

การแจกแจงปัวซง คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง ที่ค่าเฉลี่ยของความสำเร็จเกิดขึ้นในช่วงเวลาหรือขอบเขตที่กำหนด ชื่อนี้ตั้งจากนักคณิตศาสตร์ชาวฝรั่งเศสชื่อ ซีมอง เดอนี ปัวซง

เช่น จำนวนผู้ป่วยในรอบ 1 ปี (ช่วงเวลาที่กำหนดคือ 1 ปี) จำนวนเซลเม็ดเลือดขาวในเลือด 1 ลูกบาศก์มิลลิเมตร (ขอบเขตที่กำหนดคือ เลือด 1 ลูกบาศก์มิลลิเมตร) (หน้า 116)

การแจกแจงปกติ (normal distribution) คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง ที่โค้งการแจกแจงเป็นรูประฆังคว่ำสมมาตร นักคณิตศาสตร์ชาวเยอรมันชื่อ คาร์ล ฟรีดริช เกาส์ ชี้ให้เห็นการแจกแจงปกติ บางครั้งเรียกว่า การแจกแจงเกาส์เซียน (Gaussian)

รูปร่างโค้งการแจกแจงของข้อมูลของตัวแปรสุ่มชนิดต่อเนื่อง ได้จากการนำข้อมูลมาแจกแจงความถี่ ต่อไปวาดกราฟฮีสโตแกรม รูปหลายเหลี่ยมแห่งความถี่ และโค้งความถี่ตามลำดับ หากมีรูประฆังคว่ำ ลักษณะสมมาตรทั้งสองข้าง หรือเรียกว่าเส้นโค้งปกติ (หน้า 119)

คุณสมบัติของโค้งปกติ

- โค้งเป็นรูประฆังคว่ำ ลักษณะสมมาตร ที่จุด $X = \mu$
- พื้นที่ใต้โค้งแทนจำนวนข้อมูลทั้งหมด หรือค่าความน่าจะเป็น
- ปลายโค้งไม่มีโอกาสสัมผัสกับแกนนอน X ได้เลย
- ค่ามัชฌิมาเลขคณิต มัธยฐาน และฐานนิยม ยู่ตรงกัน ตรงตำแหน่งที่ $X = \mu$
- ค่าความเบ้ = 0 คือโค้งไม่เบ้ ค่าความโด่ง = 3 คือ โด่งกำลังพอดี
- ส่วนเบี่ยงเบนมาตรฐานแบ่งพื้นที่แบ่งพื้นที่ใต้โค้งปรกติออกเป็น 6 ส่วน

การแจกแจงตัวอย่าง คือการแจกแจงความน่าจะเป็นของค่าสถิติที่ได้จากตัวอย่าง คุณสมบัติของการแจกแจงค่าเฉลี่ยและค่าสัดส่วนของตัวอย่าง ได้นำไปใช้ในการอนุมานเชิงสถิติ (หน้า 128)

ค่าความคลาดเคลื่อนมาตรฐาน หมายถึงส่วนเบี่ยงเบนมาตรฐานของค่าสถิติที่ได้จากตัวอย่าง

การแจกแจงค่าเฉลี่ยของตัวอย่าง คือการแจกแจงความน่าจะเป็นของค่าเฉลี่ย($\overline{X}$) ที่สุ่มตัวอย่างจากประชากร 1 กลุ่ม การแจกแจง

การแจกแจงความแตกต่างค่าเฉลี่ยของตัวอย่าง คือ การแจกแจงความน่าจะเป็นของค่าความแตกต่างค่าเฉลี่ย ($\overline{X}_1 - \overline{X}_2$) ของตัวอย่าง ที่ได้จากการสุ่มตัวอย่างจากประชากร 2 กลุ่มที่ไม่ขึ้นต่อกัน

การแจกแจงค่าสัดส่วนของตัวอย่าง คือ การแจกแจงความน่าจะเป็นของสัดส่วน (p) ที่ที่คำนวณจากตัวอย่างขนาด n ที่สุ่มจากประชากร 1 กลุ่ม

การแจกแจงความแตกต่างค่าสัดส่วนของตัวอย่าง คือ การแจกแจงความน่าจะเป็นของความแตกต่างค่าสัดส่วน ($p_1 - p_2$) ของตัวอย่าง ที่ได้จากการสุ่มตัวอย่างจากประชากร 2 กลุ่มที่ไม่ขึ้นต่อกัน