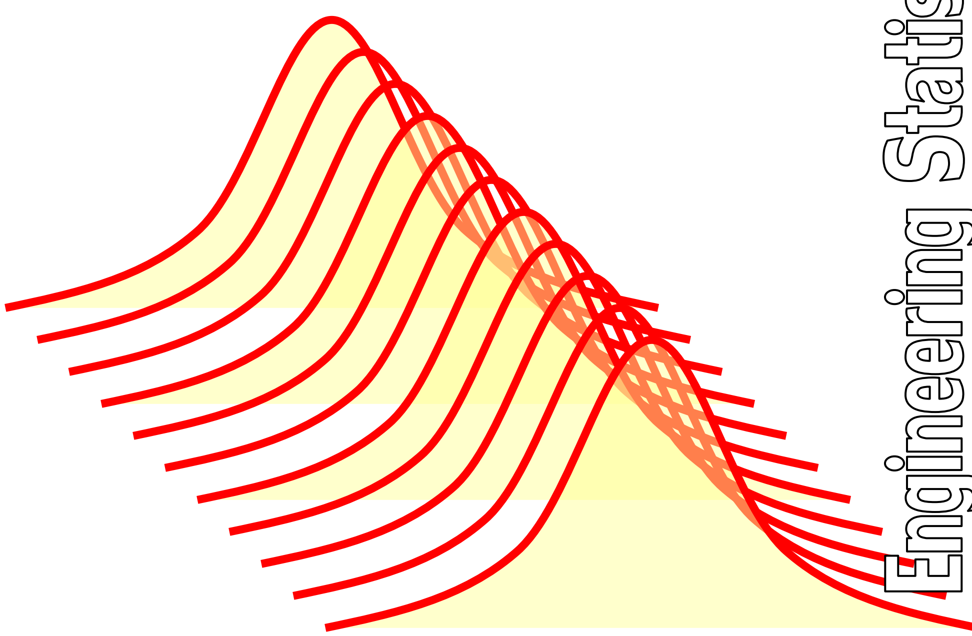
207251

สถิติทางวิศวกรรม

Engineering Statistics



รองศาสตราจารย์ ดร.วราวุธ วุฒิวณิชย์

ภาควิชาวิศวกรรมชลประทาน

คณะวิศวกรรมศาสตร์ กำแพงแสน

มหาวิทยาลัยเกษตรศาสตร์ 2547

คำนำ

สถิติทางวิศวกรรม(Engineering Statistics) เป็นวิชาที่เกี่ยวข้องกับการคำนวณความน่าจะเป็นของเหตุการณ์ต่าง ๆ โดยใช้ Tree Diagram, Venn Diagram และ Bayes' Theorem การใช้ฟังค์ชั่นการแจกแจงความน่าจะเป็นแบบต่าง ๆ อธิบายความน่าจะเป็นของตัวแปรสุ่ม และการวิเคราะห์ทางสถิติ เช่น วิธีการสุ่มตัวอย่าง การประมาณค่า การทดสอบสมมติฐาน การวิเคราะห์ความแปรปรวน และรวมถึงการวิเคราะห์การถดถอยและสหสัมพันธ์ ซึ่งถือว่าเป็นพื้นฐานที่สำคัญของวิศวกรรมศาสตร์ทุกสาขาวิชา

ผู้เขียนได้เขียนเอกสารชุดนี้ขึ้นมาจากประสบการณ์การสอนวิชาความน่าจะเป็นและสถิติประยุกต์สำหรับวิศวกรกว่า 7 ปี โดยพยายามเขียนทฤษฎีความน่าจะเป็นและหลักการวิเคราะห์ทางสถิติให้ละเอียดมากเท่าที่จะเขียนได้ และพยายามยกตัวอย่างเพื่อให้ผู้อ่านเข้าใจง่าย จึงเชื่อแน่ว่าเอกสารชุดนี้จะเป็นประโยชน์ต่อนิสิตผู้เรียนวิชานี้และวิศวกรซึ่งทำงานเกี่ยวข้องกับการวิเคราะห์ผลการทดลองซึ่งต้องใช้หลักสถิติทางวิศวกรรม

รองศาสตราจารย์ ดร.วราวุธ วุฒิวณิชย์
ภาควิชาวิศวกรรมชลประทาน
คณะวิศวกรรมศาสตร์ กำแพงแสน
มหาวิทยาลัยเกษตรศาสตร์
ตุลาคม 2547

สารบัญ

สารบัญ (ต่อ)

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

บทที่ 1
ทฤษฎีความน่าจะเป็นและสถิติเบื้องต้น (Basic Statistics and Probability Theory)

1.1 คำนำ

สถิติ (Statistics) คือ วิชาที่เกี่ยวกับการทดลองหรือการสุ่มตัวอย่าง (Experiment or Sampling) แล้วนำผลจากการทดลองหรือการสุ่มตัวอย่างมาหาข้อสรุปเกี่ยวกับประชากร (Population) โดยใช้ทฤษฎีความน่าจะเป็น (Probability) เพื่อนำไปใช้ประโยชน์เกี่ยวกับการตัดสินใจที่สำคัญ ดังรูปที่ 1.1 เช่น การพัฒนาคุณภาพการผลิตในอุตสาหกรรมต่าง ๆ ในการเกษตรและอื่น ๆ

รูปที่ 1.1 แนวความคิดเกี่ยวกับวิชาสถิติ

1.2 การทดลองทางสถิติ (Statistical Experiment)

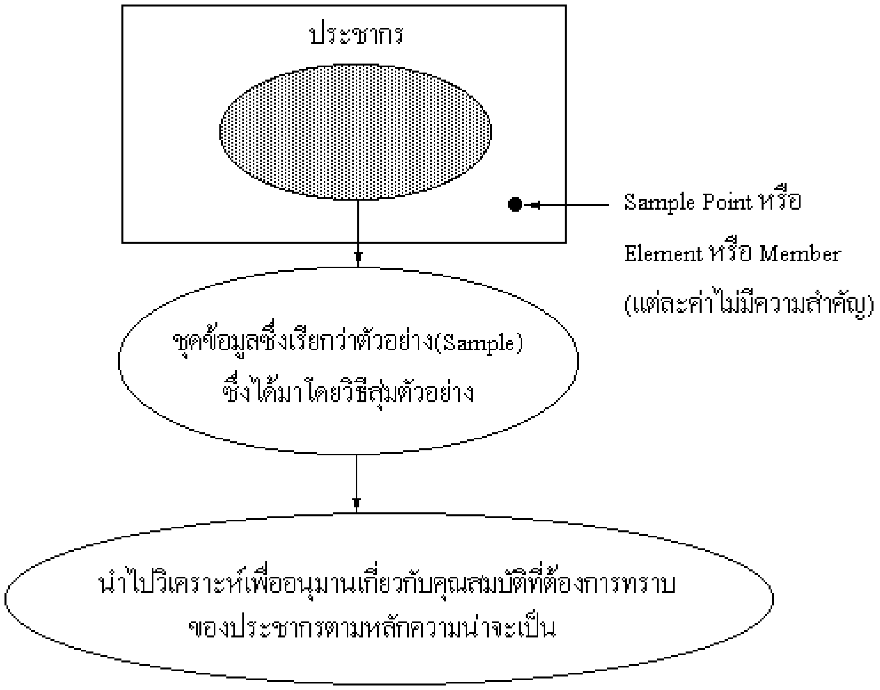
การทดลองทางสถิติ คือ กระบวนการที่ทำให้ได้มาซึ่งชุดข้อมูล (Set of Data) ที่เรียกว่าตัวอย่าง (Sample) จากประชากร เช่นการทำโพล (Poll) สุ่มตัวอย่างความคิดเห็นประชาชนเกี่ยวกับการบริหารงานของรัฐบาล ว่าเห็นด้วย หรือไม่เห็นด้วย แสดงให้เห็นถึงการสุ่มตัวอย่างและกระบวนการทางสถิติ รูปที่ 1.2

ตัวอย่างการทดลองทางสถิติและประโยชน์ของการเรียนวิชาสถิติ คือ

(1) ทราบความคิดเห็นประชาชนเกี่ยวกับการบริหารงานของรัฐบาล ซึ่งอนุมานจากโพล ***ถ้าประชาชนส่วนใหญ่ไม่เห็นด้วย รัฐบาลควรพิจารณาการปรับปรุงการบริหารงาน***

(2) ทราบสัดส่วนอุปกรณ์ที่ผลิตจากโรงงานที่ไม่มาตราฐาน ***ถ้าสัดส่วนที่ไม่มาตราฐานสูงกว่าเกณฑ์ที่โรงงานกำหนด ควรมีการปรับปรุงคุณภาพการผลิต***

(3) ทราบ Probability ของฝนรายปีที่ตกในเขต กพส. ที่มีค่าต่ำกว่า 1,000 มม. ***ถ้า Probability ของฝนดังกล่าวมีค่าสูง การเพาะปลูกโดยอาศัยน้ำฝนเพียงอย่างเดียวอาจไม่พอ ควรมีการพัฒนาการชลประทาน จึงจะทำให้ การเพาะปลูกได้ผลคุ้มค่า***



รูปที่ 1.2 กระบวนการทางสถิติ

ในการทดลองทางสถิติ มีความจริงที่สำคัญคือ ผลของการทดลองซ้ำหลาย ๆ ครั้ง อาจต่างกัน ผลของการทดลองจะขึ้นอยู่กับ Chance/Probability เช่น การทดลองครั้งที่ 2 โดยมีข้อกำหนดต่าง ๆ เหมือนเดิม อาจให้ผลลัพธ์ต่างจากการทดลองครั้งที่ 1 เนื่องจากความไม่แน่นอนต่าง ๆ (Uncertainty) ดังนั้น ทฤษฎีความน่าจะเป็นจึงมีความสำคัญในการอธิบายผลลัพธ์จากการทดลอง

1.3 ตัวแปรสุ่มและความน่าจะเป็น (Random Variable and Probability)

Random Variable (RV) = ตัวแปรสุ่มคือ ตัวแปรที่แสดงผลลัพธ์ของการทดลอง (ที่เราต้องการทราบซึ่งเป็นเลขจำนวนจริง) ไม่สามารถบอกค่าล่วงหน้าได้ แต่สามารถอธิบายค่าของ RV ได้ด้วยทฤษฎีการแจกแจงความน่าจะเป็น (Probability Distribution Function) สัญญาลักษณ์ของตัวแปรสุ่มปกติเขียนด้วยตัวพิมพ์ใหญ่ เช่น X

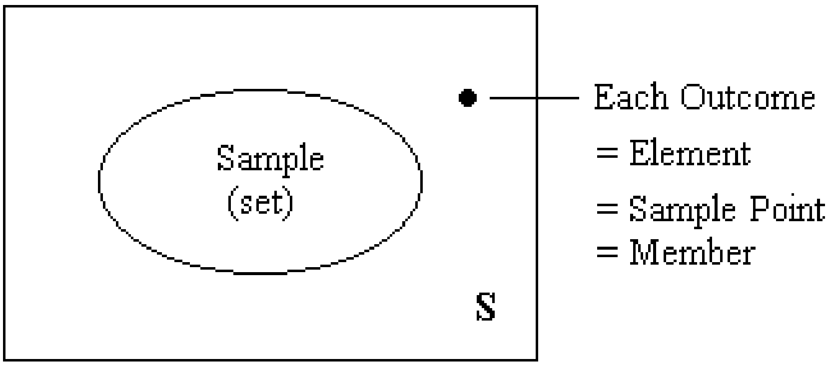
Probability = ความน่าจะเป็นของเหตุการณ์หาได้จากการทดลอง

$$f_A = \frac{n_A}{N} \quad \text{(Relative Frequency)}$$

$$P(A) = \lim_{N \to \infty} f_A$$

1.4 เอกภพสัมพัทธ์(Sample Space, S)

Sample Space คือผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลอง (Set of All Possible Outcomes of a Statistical Experiment) ดังนั้น Sample Space จึงเกี่ยวข้องกับกรทดลองอันใดอันหนึ่งโดยเฉพาะ ผลลัพธ์ของการทดลอง 1 ครั้งคือ 1 Sample Point ผลลัพธ์ของการทดลองหลาย ๆ ครั้งเรียกว่า เซท (Set) หรือตัวอย่าง (Sample) ดังแสดงในรูปที่ 1.3



รูปที่ 1.3 Sample Space

Sample Space แบ่งออกได้เป็น 2 แบบ
- นับได้ (Finite Number) หรือ Discrete Sample Space
- นับไม่ได้ (Infinite Number) หรือ continuous Sample Space

1.4.1 กรณี Sample Space เป็นจำนวนที่นับได้ (Discrete Sample Space)

การทดลองที่ 1 : โยนเหรียญ 1 เหรียญ S = {H, T}

การทดลองที่ 2 : โยนเหรียญ 2 เหรียญ S = {HH, HT, TH, TT}

การทดลองที่ 3 : ทอดลูกเต๋า 1 ลูก S1 = {1, 2, 3, 4, 5, 6}

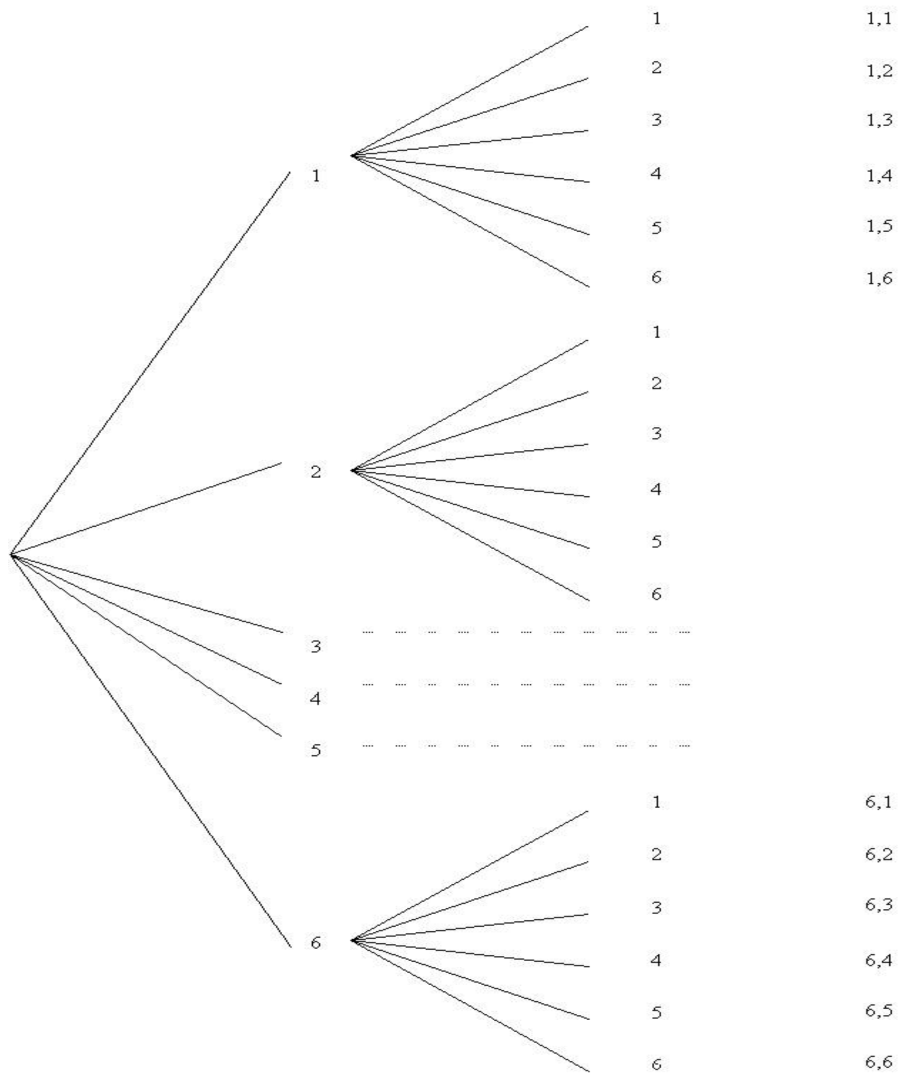
* ถ้าสนใจเฉพาะผลลัพธ์ที่เป็นเลขคู่-คี่

S2 = {คู่, คี่}

การทดลองที่ 4 : ทอดลูกเต๋า 2 ลูก

หา Sample Point ของ Sample Space โดยใช้ Tree Diagram ได้ดังนี้

First Outcome <u>ลูกที่ 1</u>	Second Outcome <u>ลูกที่ 2</u>	Sample Points



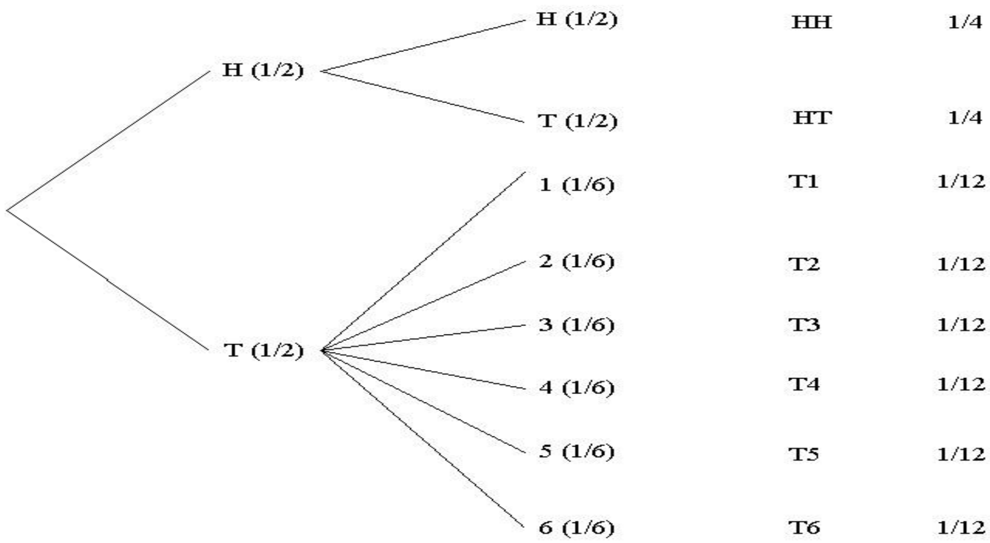
S = {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),........,(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

Tree diagram คือ เทคนิคการหา Sample Point ใน Sample Space และ Probability ของเหตุการณ์ใน S

การทดลองที่ 5 : โยนเหรียญ ถ้าออก H ให้โยนเหรียญต่ออีก 1 ครั้ง
ถ้าออก T ให้ทอดลูกเต๋า 1 ครั้ง

ให้ X คือเหตุการณ์ที่แสดงผลการโยนเหรียญ และ Y คือ เหตุการณ์ที่แสดงผลการทอดลูกเต๋า

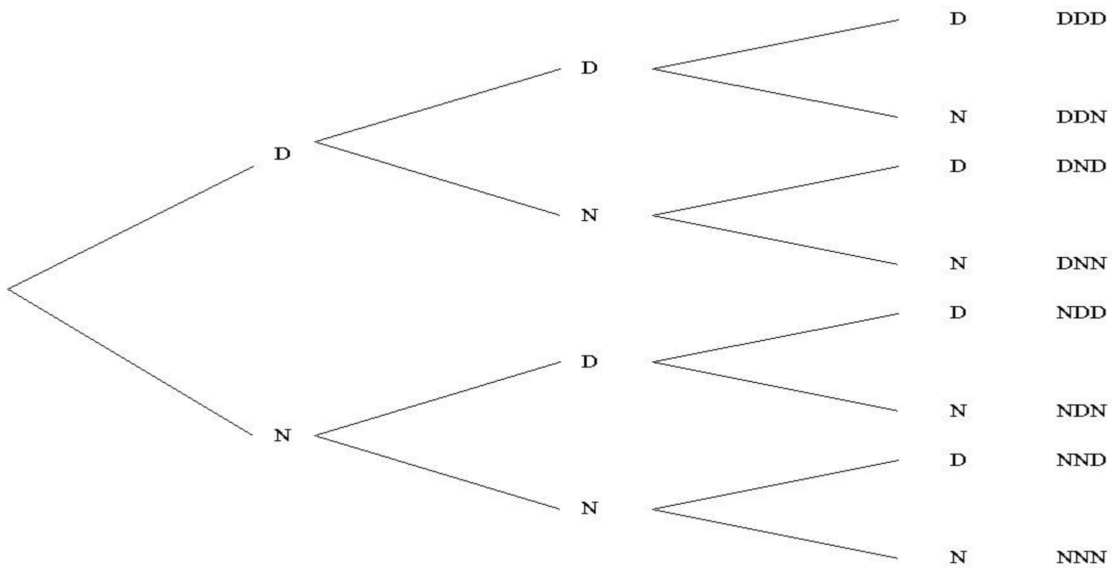
First Outcome	Second Outcome	Sample Point	P(X,Y)
X;P(X)	Y;P(Y|X)	(X,Y)	



S = {HH, HT, T1, T2, T3, T4, T5, T6}

การทดลองที่ 6 : สุ่มตัวอย่างผลิตภัณฑ์ 3 ชิ้น ในการสุ่มแต่ละครั้งได้ผล 2 แบบ คือ D (defective), N = (Non-defective)

สุ่มครั้งที่ 1 สุ่มครั้งที่ 2 สุ่มครั้งที่ 3 Sample Point



S = {DDD, DDN, DND, DNN, NDD, NDN, NND, NNN}

หมายเหตุ ถ้าทราบ Sample Space จะสามารถหาค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม และการแจกแจงความน่าจะเป็นของตัวแปรสุ่มนั้นได้ เช่น จากการทดลองที่ 6 ถ้าให้ X คือตัวแปรสุ่มแสดงจำนวนผลิตภัณฑ์เสียและความน่าจะเป็นที่จะสุ่มได้ผลิตภัณฑ์เสียเท่ากับ ½ จะได้การแจกแจงความน่าจะเป็นของ X นี้

x	0	1	2	3
P(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

1.4.2 กรณี Sample Space เป็นจำนวนที่นับไม่ได้ (Continuous Sample Space)

กรณีนี้จะอธิบายด้วย Statement หรือ Rule เช่น ผลลัพธ์ของการทดลองคือ

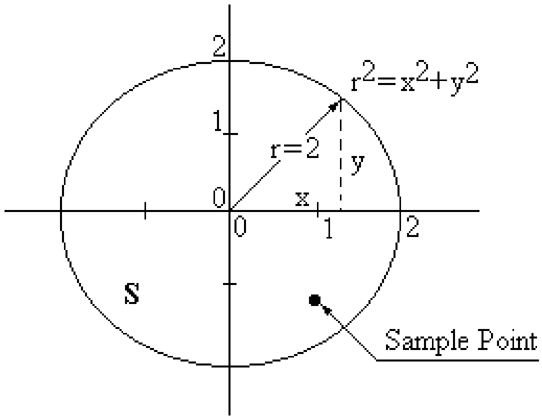
S = {Set of cities ในโลกที่มีพลเมืองมากกว่า 1 ล้านคน}

= {x such that x is a city with a population > 1 million}

หรือ S = $\{(x,y) \mid x^2 + y^2 \le 4\}$

= {points x, y ในวงกลมซึ่งมี radius 2 และจุดศูนย์กลางอยู่ที่ origin}
(ดูรูปที่ 1.4)

หรือ S = $\{t \mid t \geq 0\}$ t = อายุการใช้งานของอุปกรณ์เป็นปี



รูปที่ 1.4 Sample Space { (x,y) / $x^2 + y^2 \leq 4\}$

1.5 เหตุการณ์ (Event)

เหตุการณ์คือ เซทย่อยของ Sample Space หรือเซทที่สนใจ ในการทดลอง เราไม่ได้สนใจ Sample Point อันใดอันหนึ่งโดยเฉพาะ โอกาสในการเกิด Sample Point อันใดอันหนึ่งมีค่าน้อยมากใน Sample Space ขนาดใหญ่ เช่น

ให้ X = ฝนรายปี ที่ กพส. เป็นตัวแปรสุ่ม

P (X = 1,000 มม/ปี) $\longrightarrow 0$

แต่เราจะสนใจการเกิดเหตุการณ์ที่สำคัญ เช่น

A = {X ≤ 1,000 มม/ปี}

A = เหตุการณ์ที่ฝนในปีใด ๆ น้อยกว่าหรือเท่ากับ 1,000 มม. ซึ่งถือว่าฝนน้อย

ถ้า P(A) มีค่ามาก แสดงว่าการเกษตรน้ำฝนจะไม่ประสบผลต้องมีการชลประทานช่วย

หรือ B = $\{t \mid 0 \leq t < 5\}$

เมื่อ t = อายุการใช้งานของอุปกรณ์ (ปี)

B = เหตุการณ์ที่อุปกรณ์จะเสียก่อนสิ้นปีที่ 5

1.5.1 Complement of A (A')

เหตุการณ์ที่ไม่อยู่ใน A (Event that is not in A.)

1.5.2 Intersection of A and B ($A \cap B$)

$A \cap B$ คือเหตุการณ์ซึ่งทุกค่าอยู่ทั้งใน A และ B

(Event containing all sample points that are common to A and B)

Mutually Exclusive เหตุการณ์ที่ไม่เกิดร่วมกัน

(Events that have no elements in common)

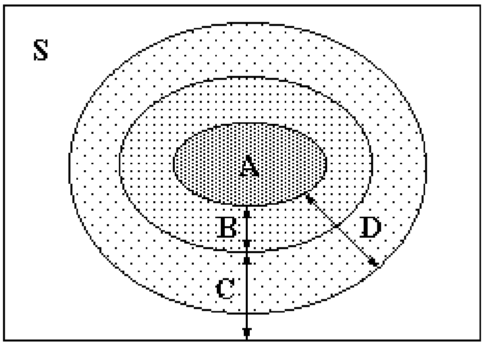
$A \cap B = \varnothing$ ถ้า A, B are mutually exclusive events

1.5.3 Union of A and B ($A \cup B$)

$A \cap B$ = เหตุการณ์ที่ทุกค่าอยู่ใน A หรือ B หรือทั้ง A และ B

(Event containing all elements that belong to A or B or both.)

1.6 สรุปกฎเกี่ยวกับ Set



ให้ X = ฝนรายปีในปีใดที่ กพส. มีค่าอยู่ระหว่าง 0-∞

S = {0-∞}

A = {ฝน <1,000 มม./ปี : ฝนน้อย}

B = {ฝน 1,000-1,300 มม./ปี : ฝนปกติ}

C = {ฝน >1,300 มม./ปี: ฝนมาก}

D = {ฝน 1,000-1,600 มม./ปี : ฝนปกติ-ค่อนข้างมาก}

รูปที่ 1.5 ไดอะแกรมแสดง Sample Space และเหตุการณ์ที่เกี่ยวกับการตกของฝนในปีใด

ถ้า A, B, C เป็นเหตุการณ์ที่เราสนใจ (Event)

(1) Union [U]

AUB = {ฝน $\leq$ 1,300 มม./ปี : ฝนน้อย-ปกติ}

(2) Intersection ($\cap$)

$B \cap D$ = B = {ฝน 1,000-1,300 มม./ปี}

(3) Complement

A' = S-A = {ฝน 1,000-∞ มม./ปี}

(4) Mutually Exclusive คือ เหตุการณ์ที่ไม่เกิดร่วมกัน ถ้าเกิดเหตุการณ์หนึ่งอีกเหตุการณ์จะไม่เกิดเช่น A กับ A'

(5) กฎการคำนวณเกี่ยวกับ Set ที่ควรจำ

1. AUB = BUA

$A \cap B = B \cap A$

2. $(A \cup B) \cup C = A \cup (B \cup C)$

$(A \cap B) \cap C = A \cap (B \cap C)$

3. $A \cap (B \cap C) = (A \cap B) \cap (A \cap C)$

$A \cup (B \cup C) = (A \cup B) \cup (A \cup C)$

4. $A \cap S = A$

$A \cup S = S$

5. $A \cap A' = \varnothing$ (null set)

$A \cup A' = S$

$A \cap \varnothing = \varnothing$

$A \cup \varnothing = A$

6. $S' = \varnothing$

$\varnothing' = S$

$(A')' = A$

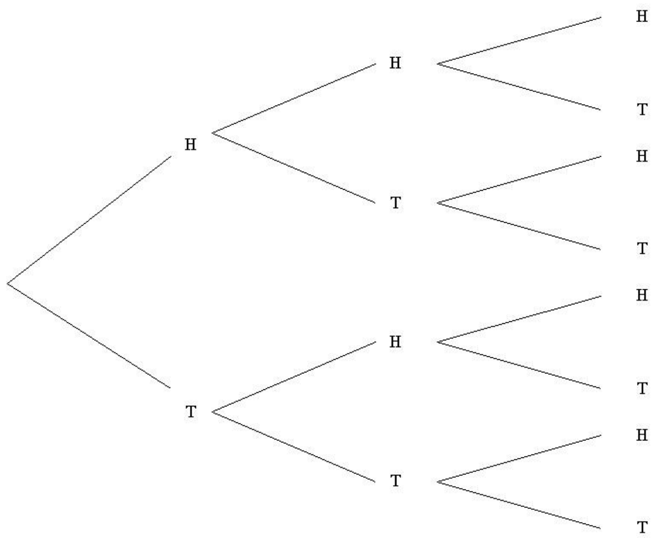
7. $(A \cap B)' = A' \cup B'$

$(A \cup B)' = A' \cap B'$

8. $(A \cap B) \cup (A \cap B') = A$

1.7 ไดอะแกรมต้นไม้ (Tree diagram) คือ เทคนิคในการหา Sample Points ใน Sample Space และหาค่าความน่าจะเป็นของเหตุการณ์ที่ซับซ้อน

เช่น โยนเหรียญ 3 ครั้ง ให้ H = ออกหัว, T = ออกก้อย

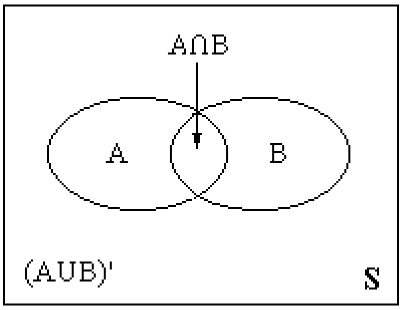


S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

$P(\text{ไม่ออกหัวเลย}) = \frac{1}{8}$

$P(\text{ออกหัว 2 ครั้ง}) = \frac{3}{8}$

1.8 เวนน์ไดอะแกรม (Venn Diagram) คือ เทคนิคในการอธิบายเรื่อง Set หรือ Event โดยวิธีกราฟฟิกแบบง่าย มีประโยชน์ต่อการคำนวณหาค่าความน่าจะเป็นของเหตุการณ์ที่ซับซ้อน (Compound Event) ดังรูปที่ 1.6



$S = (AUB) \cup (AUB)'$

$$P(A) = \frac{\text{No. of Sample Points in A}}{\text{No. of Sample Points in S}} = \frac{A}{S}$$

$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

เมื่อ P(B|A) คือ Conditional Probability หรือความน่าจะเป็นแบบมีเงื่อนไขของเหตุการณ์ B ซึ่งเกิดใน A

รูปที่ 1.6 Venn Diagram

1.9 กฎการนับจำนวนเหตุการณ์ (Counting Sample Point หรือ Basic Counting Rule)

กฎการนับจำนวนเหตุการณ์ คือวิธีการคำนวณหา No. of Sample Points ใน Sample Space โดยไม่ต้องการทราบว่าแต่ละ Sample Point คืออะไร

1.9.1 Tree Diagram (ดูข้อ 1.8)/Venn Diagram (ดูข้อ 1.7)

1.9.2 Multiplicative Rule

เลือกของ k อย่างจากของ k ชุด (sets) ซึ่งแตกต่างกัน แต่ละชุดมีจำนวนเท่ากับ $n_1, n_2, \ldots\ldots, n_k$ ตามลำดับ

No. of Sample Points $= n_1 . n_2 \ldots\ldots\ldots n_k$ เช่น

(1) การทอดลูกเต๋า 2 ลูกพร้อมกันจะได้ผลลัพธ์กี่แบบ

ลูกที่ 1 : n_1 = 6

ลูกที่ 2 : n_2 = 6

No. of Sample Points $= n_1 \times n_2$

$= 6 \times 6$

$= 36$

(2) Soups = 4 อย่าง (n_1)

Sandwiches = 3 อย่าง (n_2)

Desserts = 5 อย่าง (n_3)

Drinks = 4 อย่าง (n_4)

จัดชุดอาหารให้มีครบทุกชนิด ได้กี่แบบ

No. of Sample Points

$= n_1 \times n_2 \times n_3 \times n_4$

$= 4 \times 3 \times 5 \times 4$

$= 240$

1.9.3 การจัดลำดับ (Permutation)

- arrangement of all or part of a set of objects.

(1) การจัดลำดับหรือการเลือกของ r สิ่งจากของทั้งหมด n สิ่งซึ่งแตกต่างกัน (Distinct Objects) โดยคำนึงถึงลำดับ เช่น ABCD ≠ DABC

$$nPr = P(n,r) = \frac{n!}{(n-r)!}$$

เช่น จับ Lottery 2 รางวัล (รางวัลที่ 1 และ 2) จาก Lottery ทั้งหมด 20 ใบ จงหา No. of Sample Points

$n = 20$

$r = 2$

$$20P2 = P(20,2) = \frac{20!}{(20-2)!} = \frac{20!}{18!} = 20 \times 19$$

$= 880$

* เลือกรางวัลที่ 1 จาก Lottery 20 ใบ รางวัลที่ 2 จะเลือกจาก 19 ใบที่เหลือ

(2) การจัดลำดับของ n สิ่ง จากของทั้งหมด n สิ่งซึ่งต่างกัน

$$nPn = P(n,n) = n!$$

เช่น มี Soup 1 ที่

Sandwich 1 ที่

Dessert 1 ที่

Drink 1 ที่

สามารถจัดลำดับได้กี่แบบ

$n = 4,\ r = n = 4$

$4P4 = P(4,4) = 4! = 4 \times 3 \times 2 \times 1 = 24$

* Permutation ต่างจาก Multiplicative Rule ตรงที่การทดลองแต่ละครั้งไม่อิสระต่อกัน เช่น ถ้าครั้งแรกเลือก Soup อาหารที่เหลือจะไม่มี Soup เป็นต้น

(3) Circular Permutation

การจัดลำดับของ n สิ่งซึ่งต่างกันเป็นวงกลม

$CP(n,n) = (n-1)!$

ในการจัดของ n สิ่งซึ่งแตกต่างกัน (Distinct Object) เป็นวงกลม ถ้าจัดหมุนตามเข็มนาฬิกา 1 ตำแหน่ง การจัดลำดับจะยังคงเหมือนเดิม ดังนั้นวิธีคิดคือ Fixed ตำแหน่งของ 1 สิ่งแล้วจัดของที่เหลือ (n-1) สิ่ง

หรือพิจารณาว่า Permutation จะลดลงเท่ากับจำนวนครั้งที่หมุนตามเข็มนาฬิกาจนครบรอบ ซึ่งเท่ากับ n

ตัวอย่าง จัดนักเรียน 5 คน เข้าแถว 5 ที่

$P(5,5) = 5! = 120$

ถ้าจัดเป็นวงกลม

$CP(5,5) = 4! = 24$

(4) การจัดลำดับของ n สิ่งซึ่งไม่ต่างกันทั้งหมด

มีของชนิดที่ 1 = n_1

มีของชนิดที่ 2 = n_2

. .

มีของชนิดที่ k = n_k

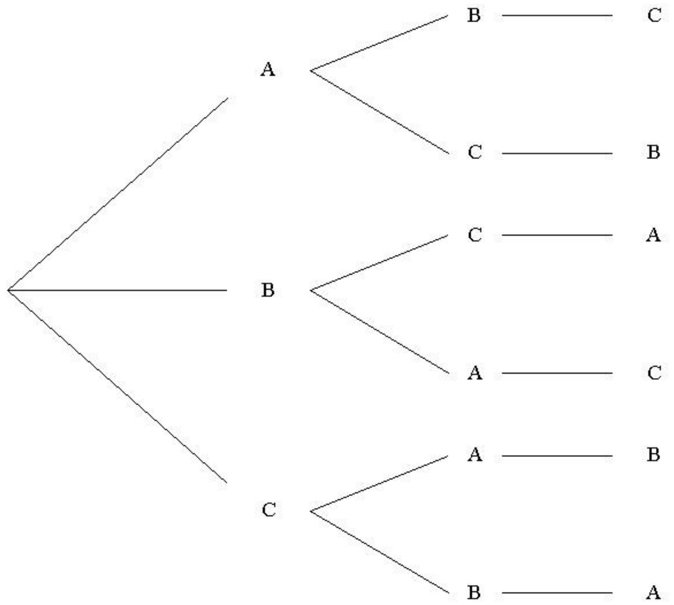
$n = n_1 + n_2 + \ldots\ldots\ldots\ldots + n_k$

No. of Permutation = $\dfrac{n!}{n1!\,n2!\,\ldots\ldots nk!}$

(จำนวนแบบของการจัดลำดับ)

$= \binom{n}{n1,\ n2, \ldots\ldots,\ nk}$

ตัวอย่างที่ 1.1 จงหาวิธีการจัดเรียงตัวอักษร A, B, C (n=3)

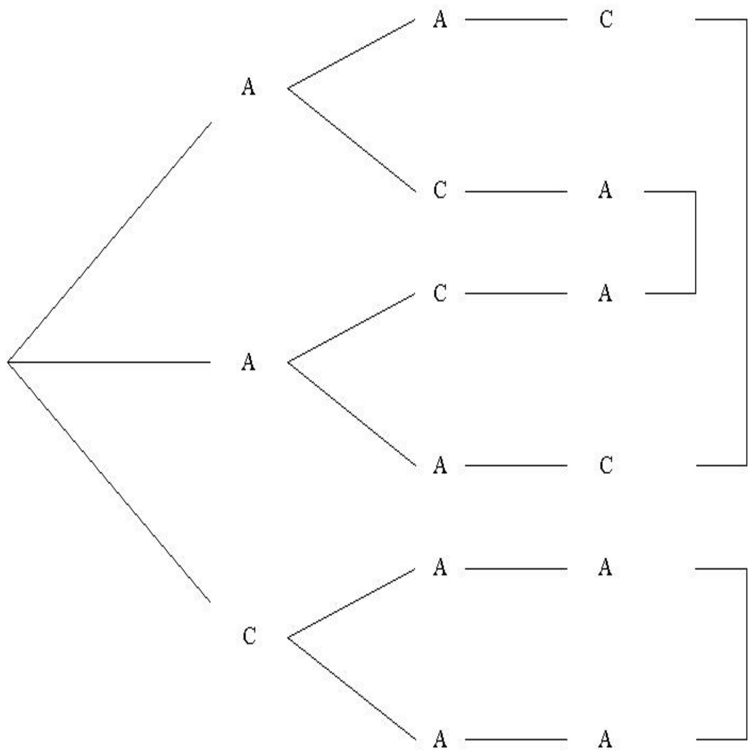


No. of Permutation $= 3!$

$= 3 \times 2 \times 1 = 6$

ถ้า B = A จงหาวิธีการจัดเรียงตัวอักษร A, A, C

($n=3$; $n_1=2$, $n_2=1$)



No. of Permutation

$$= \frac{n!}{n1!n2!}$$

$$= \frac{3!}{2!\,1!}$$

$$= 3$$

จำนวนลดลงเป็นสัดส่วนกับจำนวน Permutation ของที่เหมือนกัน กรณีนี้เท่ากับ 2!

1.9.4 การแบ่งกลุ่ม (Partitioning)

แบ่งของ n สิ่งเป็น r กลุ่ม

กลุ่มแรกมี n_1 สิ่ง

กลุ่มสองมี n_2 สิ่ง

. . .

กลุ่ม r มี n_r สิ่ง

$$n_1 + n_2 + \dots + n_r = n$$

$$\binom{n}{n1,\ n2, \dots,\ nr} = \binom{n}{r} = \frac{n\,!}{n1\,!\ n2\,! \dots nr\,!}$$

* สูตรเหมือนกรณี Permutation ของ n สิ่งซึ่งไม่ต่างกันทั้งหมด

1.9.5 การจัดหมู่ (Combinations)

จัดของ r สิ่งจากของ n สิ่งที่แตกต่างกันโดยไม่คำนึงถึงลำดับ เช่น ABCD = DABC

$$nCr = C(n,r) = \frac{n\,!}{(n-r)\,!\ r\,!}$$

1.9.6 การสุ่มหรือการจัดลำดับกรณีมีข้อแม้เพิ่ม

1.9.6.1 ใช้วิธีผสม Combination + Multiplicative

ตัวอย่างที่ 1.2 จากปัญหาข้อ 1.9.2 (2) ให้จัดชุดอาหารครบทุกชนิด ชนิดละ 2 อย่าง (แตกต่างกัน) สามารถจัดได้กี่แบบ (ไม่คำนึงถึงลำดับ)

จัด Soup ได้ $= C(4,2) = \frac{4!}{2!\,2!} = \frac{4x3}{2} = 6$

จัด Sandwiches ได้ $= C(3,2) = \frac{3!}{1!\,2!} = 3$

จัด Desserts ได้ $= C(5,2) = \frac{5!}{3!\,2!} = \frac{5x4}{2} = 10$

จัด Drinks ได้ $= C(4,2) = \frac{4!}{2!\,2!} = 6$

$\therefore$ จำนวนชุดอาหาร $= 6x3x10x6$

$= 1{,}080$

ตัวอย่างที่ 1.3 หลอดไฟ 10 ดวง เสีย 3 ดวง ถ้าต้องการเลือกหลอดไฟ 3 ดวงโดยไม่คำนึงถึงลำดับจะเลือกได้กี่วิธี โดยมีข้อแม้ดังนี้

S = {หลอดเสีย 0, หลอดเสีย 1, หลอดเสีย 2, หลอดเสีย 3}

(1) ไม่มีข้อแม้

จำนวนวิธี = $C(10,3) = \frac{10!}{7!\,3!} = \frac{10x9x8}{3x2} = 120$

(2) ไม่ได้หลอดเสียเลย (หลอดเสีย 0 หลอดดี 3)

หยิบ 3 หลอดจากหลอดดีทั้งหมด 7 หลอด

จำนวนวิธี = $C(7,3) x\, C(3,0)$

$= \frac{7!}{4!\,3!} x\, 1 = \frac{7x5x6}{3x2} = 35$

(3) ได้หลอดเสีย 1 (หลอดดี 2)

หยิบ 1 หลอด จากหลอดเสีย 3

หยิบ 2 หลอด จากหลอดดี 7

จำนวนวิธี = $C(7,2)\, C(3,1)$

$= \frac{7!}{5!\,2!} x \frac{3!}{2!\,1!} = 21x3 = 63$

(4) ได้หลอดเสีย 2 (หลอดดี 1)

จำนวนวิธี = $C(7,1)\, C(3,2)$

$= \frac{7!}{6!\,1!} x \frac{3!}{2!\,1!} = 7x3 = 21$

(5) ได้หลอดเสีย ≤ 2

จำนวนวิธี = 35+63+21 = 119

(6) ได้หลอดเสีย 3

จำนวนวิธี = $C(7,0)\, C(3,3)$

$= \frac{7!}{7!\,0!} x \frac{3!}{0!\,3!} = 1$

หรือ = 120-119 = 1

ตัวอย่างที่ 1.4 เลือกตัวอักษร 4 ตัว จากคำว่า EXPRESSION ได้กี่วิธี

E = 2; S = 2; X, P, R, I, O, N อย่างละ 1 ตัว

S = {ตัวอักษรเหมือนกัน 2 ตัว 2 คู่, เหมือน 2 ตัว ต่าง 2 ตัว, ต่างกันทั้ง 4 ตัว}

วิธีที่ 1 ตัวอักษรเหมือนกัน 2 ตัว 2 คู่

จำนวนวิธี = $C(2,2)\, C(7,0) = 1$

วิธีที่ 2 ตัวอักษรเหมือนกัน 2 ตัว ต่าง 2 ตัว

จำนวนวิธี = $C(2,1)\, C(7,2) = 2x\frac{7x6}{2} = 42$

วิธีที่ 3 ตัวอักษรต่างกันทั้งหมด (หยิบ E 2 ตัว หรือ S 2 ตัวไม่ได้)

จำนวนวิธี = $C(8,4) = \frac{8x7x6x5}{4x3x2x1} = 70$

∴ จำนวนวิธีทั้งหมด = 1+42+70 = 113

1.9.6.2 ใช้วิธีผสม Permutation + Multiplicative

เช่น จากปัญหาข้อ 1.9.3(3) จัดนักเรียน 5 คนเข้าแถว 5 ที่ ถ้านักเรียน 2 คน ต้องยืนชิดกัน

(--)--- n = 4

P(4,4) = 4!

(--) n = 2

P(2,2) = 2!

∴ จัดได้ = 4! X 2! = 48 แบบ

ถ้า 2 คนไม่ยอมยืนชิดกัน

จัดได้ = 5! - 48 = 120-48 = 72 แบบ

1.9.6.3 ใช้วิธีผสม Combination + Permutation

จากตัวอย่างในหัวข้อ 1.9.6.2 ให้หาวิธีจัดลำดับ (Permutation) ตัวอักษร 4 ตัวที่เลือกจากคำว่า EXPRESSION

วิธีที่ 1 เหมือน 2 ตัว 2 คู่

$n = 4$; $n_1 = 2$; $n_2 = 2$

No. of Permutation $= \dfrac{4!}{2!\,2!} = 6$

วิธีที่ 2 เหมือน 2 ต่าง 2

จำนวนวิธีเลือก (No. of Combination) = C(2,1) C(7,2) = 42

จำนวนวิธีจัดเรียงหลังจากเลือก 4 ตัว (No. of Permutation)

$$= \frac{4!}{2!\,1!\,1!} = 12$$

∴ จำนวนวิธีทั้งหมด = 42 x 21 = 504

วิธีที่ 3 ตัวอักษรต่างกันทั้งหมด

No. of Combination = C(8,4) = 70

No. of Permutation = P(4,4) = 4! = 24

∴ จำนวนวิธีทั้งหมด = 70 x 24 = 1,680

รวมจำนวนวิธีเลือกทุกวิธี= 6+504+1,680

= 2,190

1.10 ความน่าจะเป็น (Probability)

$0 \leq P(A) \leq 1$

$P(S) = 1$

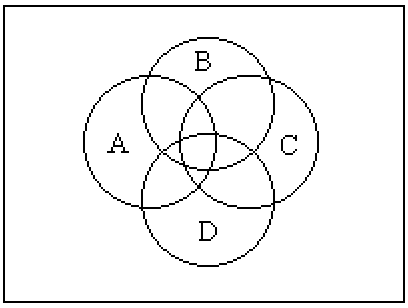
$P(\varnothing) = 0$

1.10.1 Additive rule (Addition rule)

$P(A \cup B) = P(A)+P(B)-P(A\cap B)$

$P(A \cup B \cup C) = P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$
$-P(B\cap C)+P(A\cap B\cap C)$

จากรูป $P(A \cup B \cup C \cup D) = ?$



$P(A\cup B\cup C\cup D) = P(A)+P(B)+P(C)+P(D)-P(A\cap B)-P(B\cap C)$
$-P(C\cap D)-P(D\cap A)-P(A\cap C)+P(A\cap B\cap C)$
$+P(A\cap C\cap D)+P(A\cap B\cap D)-P(A\cap B\cap C\cap D)$

* Mutually exclusive event

$P(A\cap B) = \varnothing$ if A and B are mutually exclusive

Thus $P(A \cup B) = P(A)+P(B)$

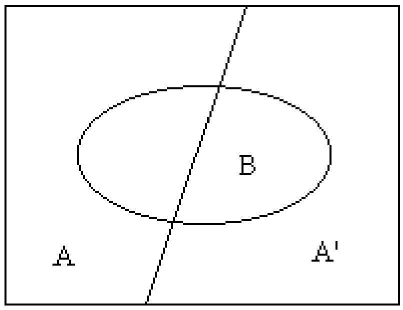
1.10.2 Multiplicative rule (Multiplication rule)

$P(A\cap B) = P(B)P(A|B) = P(A)P(B|A)$

$P(A\cap B\cap C) = P(A)P(B|A)P(C|A\cap B)$
$= P(B)P(C|B)P(A|B\cap C)$
$= P(C)P(A|B)P(B|C\cap A)$

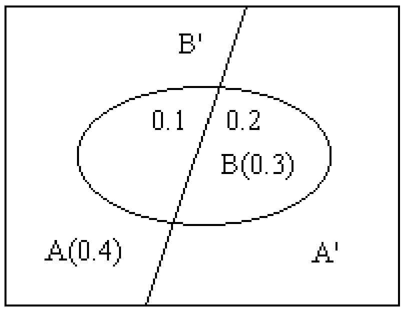
* $P(B|A) = P(B)$ if A and B are independent

Thus $P(A\cap B) = P(A)P(B)$



*** Conditional Probability, P (B|A)**

คือ ความน่าจะเป็นแบบมีเงื่อนไขของเหตุการณ์ B ที่เกิดใน A เช่น



กำหนดให้

$P(A)$	=	0.4		
$P(B)$	=	0.3		
$P(A \cap B)$	=	$P(B \cap A)$	=	0.1
$P(A' \cap B)$	=	$P(B \cap A')$	=	0.2

วิธีการคำนวณ Probability แบบต่าง ๆ จาก Venn Diagram

$$P(A') = 1-0.4 = 0.6$$

$$P(B') = 1-0.3 = 0.7$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.1}{0.4} = 0.25$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.3} = 0.333$$

$$P(B|A') = \frac{P(A' \cap B)}{P(A')} = \frac{0.2}{0.6} = 0.333$$

$$P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{0.2}{0.3} = 0.667$$

$$P(A \cap B') = P(A) - P(B \cap A)$$

$$= 0.4 - 0.1 = 0.3$$

$$P(A'\cap B') = P(A')-P(B\cap A')$$
$$= 0.6 - 0.2 \quad = 0.4$$

$$P(B'|A) = \frac{P(A\cap B')}{P(A)} = \frac{0.3}{0.4} = 0.75$$

$$P(A|B') = \frac{P(A\cap B')}{P(B)} = \frac{0.3}{0.7} = 0.4289$$

$$P(B'|A') = \frac{P(A'\cap B')}{P(A')} = \frac{0.4}{0.6} = 0.667$$

$$P(A'|B') = \frac{P(A'\cap B')}{P(B')} = \frac{0.4}{0.7} = 0.571$$

1.10.3 Total probability rule

For any two events A and B

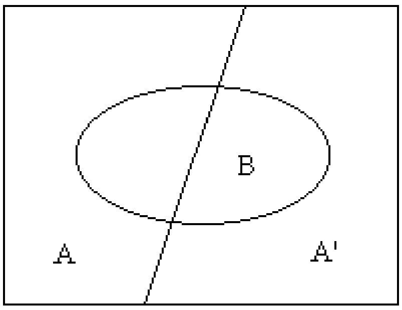
$$P(B) = P(B\cap A)+P(B\cap A')$$
$$= P(A)P(B|A)+P(A')P(B|A')$$

1.11 ทฤษฎีของ Bayes (Bayes' Theorem)

ใช้หา Conditional Probability ของเหตุการณ์

กำหนดให้ $P(A)$, $P(B|A)$, $P(B|A')$

ถ้า A,A' เป็น Mutually Exclusive Event และ P(A) หรือ P(A') ไม่เท่ากับ 0



$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$
$$= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$$

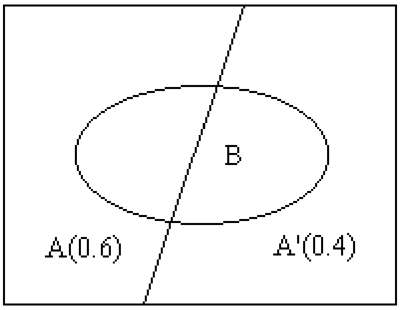
เช่น ในการสร้างทางหลวงแห่งหนึ่งให้

A = เหตุการณ์ที่มีการสไตร๊ค์

B = เหตุการณ์ที่งานเสร็จตามกำหนด

P(A) = 0.6 ; P(B|A) = 0.35

P(A') = 0.4 ; P(B|A') = 0.85



Venn Diagram

Tree Diagram

X	Y|X	X,Y	P(X,Y)	P(Y)
	B (0.35)	$A\cap B$	0.21	
A				P(B)=0.55
(0.6)	B'(0.65)	$A\cap B'$	0.39	
	B (0.85)	$A'\cap B$	0.34	
A'				P(B')=0.45
(0.4)	B'(0.15)	$A'\cap B'$	0.06	

X = เหตุการณ์แสดงสภาพของการสร้างทาง (A,A')

Y = เหตุการณ์แสดงผลของการสร้างทาง (B,B')

		Y		P(X)
		B	B'	Marginal
X	A	0.21	0.39	0.6
	A'	0.34	0.06	0.4
P(Y) Marginal		0.55	0.45	1.00

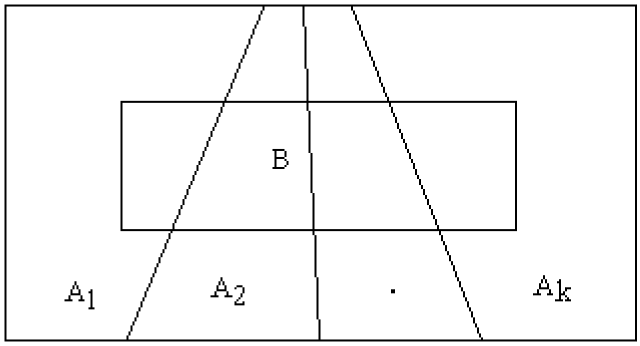
$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

$$= \frac{0.21}{0.55} = 0.38$$

$$P(A'|B) = \frac{0.34}{0.55} = 0.62$$

ถ้า $P(X).P(Y) = P(X,Y)$ แสดงว่า X,Y independent

ถ้า $P(X).P(Y) \neq P(X,Y)$ แสดงว่า X,Y ไม่ independent (Conditional)



ถ้า $A_1, A_2, ...A_k$ เป็น k Mutually Exclusive Events และ $P(A;) \neq 0$

$$P(B) = \sum_{i=1}^{k} P(A_i)P(B|A_i)$$

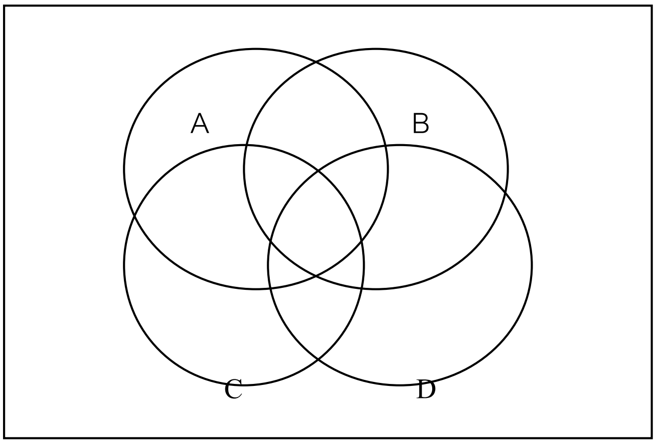
$$P(A_r \mid B) = \frac{P(A_r)\, P(B \mid Ar)}{P(B)}$$

1.12 แบบฝึกหัด

1. กล่องใบหนึ่งมีลูกบอล 8 ลูก สีแดง 3 สีเขียว 2 และสีขาว 3 จงคำนวณหาวิธีการหยิบลูกบอลจากเงื่อนไขต่าง ๆ ดังต่อไปนี้ พร้อมเขียน Tree Diagram หรือแสดงวิธีการเขียน Tree Diagram กรณีที่จำนวนวิธีมาก

เงื่อนไข	เงื่อนไขลูกบอลที่หยิบ	คำนึงถึงลำดับ	ข้อพิจารณาเกี่ยวกับลูกบอล
1.1	3 ลูกใด ๆ	ใช่	ทั้ง 8 ลูกต่างกัน
1.2	3 ลูกใด ๆ	ไม่	ทั้ง 8 ลูกต่างกัน
1.3	3 ลูกใด ๆ	ใช่	บอลสีเดียวกันไม่ต่างกัน
1.4	3 ลูกใด ๆ	ไม่	บอลสีเดียวกันไม่ต่างกัน
1.5	3 ลูก สีละลูก	ใช่	ทั้ง 8 ลูกต่างกัน
1.6	3 ลูก สีละลูก	ไม่	ทั้ง 8 ลูกต่างกัน
1.7	6 ลูก สีละ 2 ลูก	ใช่	ทั้ง 8 ลูกต่างกัน
1.8	6 ลูก สีละ 2 ลูก	ไม่	ทั้ง 8 ลูกต่างกัน
1.9	8 ลูก	ใช่	ทั้ง 8 ลูกต่างกัน
1.10	8 ลูก	ไม่	ทั้ง 8 ลูกต่างกัน
1.11	8 ลูก	ใช่	บอลสีเดียวกันไม่ต่างกัน

2. จงหา P(AUBUCUD) จาก Venn Diagram ดังรูป



3. มีตู้ 3 ตู้ (A, B, C) แต่ละตู้มีลิ้นชัก 2 อัน แต่ละลิ้นชักใส่เหรียญไว้ 1 เหรียญ ดังนี้

ตู้	เหรียญ
A	2 เหรียญทอง
B	2 เหรียญเงิน
C	1 เหรียญทอง 1 เหรียญเงิน

สุ่มเปิดลิ้นชักแบบไม่เจาะจง ถ้าพบเหรียญเงิน จงหาความน่าจะเป็นที่อีกลิ้นชักหนึ่งของตู้นี้มีเหรียญเงินบรรจุอยู่ (เขียน Tree Diagram ประกอบ)

4. มีกล่อง 2 ใบ (A และ B) แต่ละใบบรรจุลูกบอล 9 ลูก ดังนี้

กล่อง	ลูกบอล
A	สีแดง 4 ลูก สีฟ้า 3 ลูก สีเขียว 2 ลูก
B	สีแดง 2 ลูก สีฟ้า 3 ลูก สีเขียว 4 ลูก

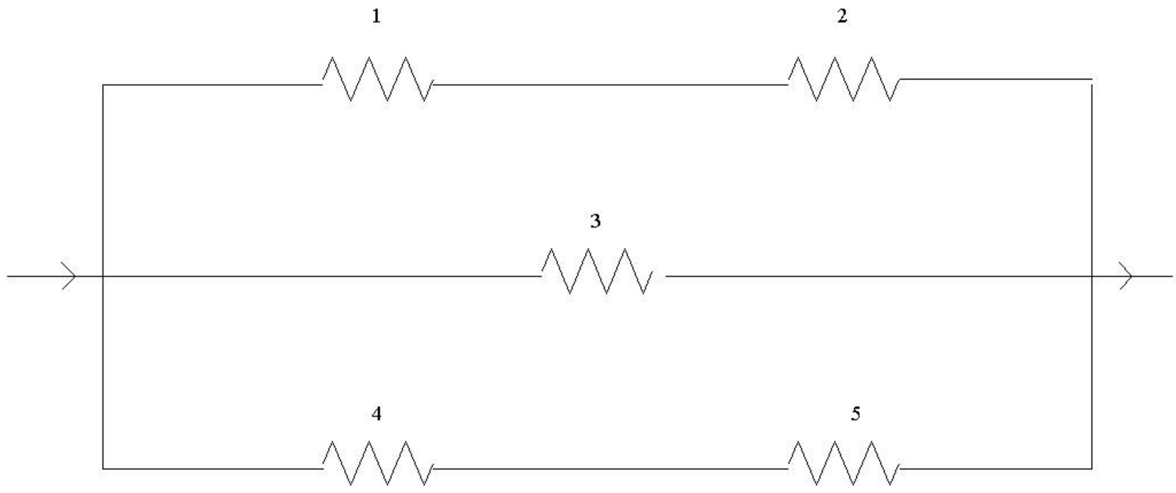
ถ้าหยิบลูกบอล 1 ลูกจากกล่อง A แล้วใส่ลงในกล่อง B แล้วจึงหยิบลูกบอล 1 ลูกจากกล่อง B

จงหา (ก) P (ลูกบอลที่หยิบจากกล่อง B เป็นสีแดง)

(ข) ถ้าลูกบอลที่หยิบจากกล่อง B เป็นสีแดง จงหา P (ลูกบอลที่หยิบจากกล่อง A เป็นสีแดง)

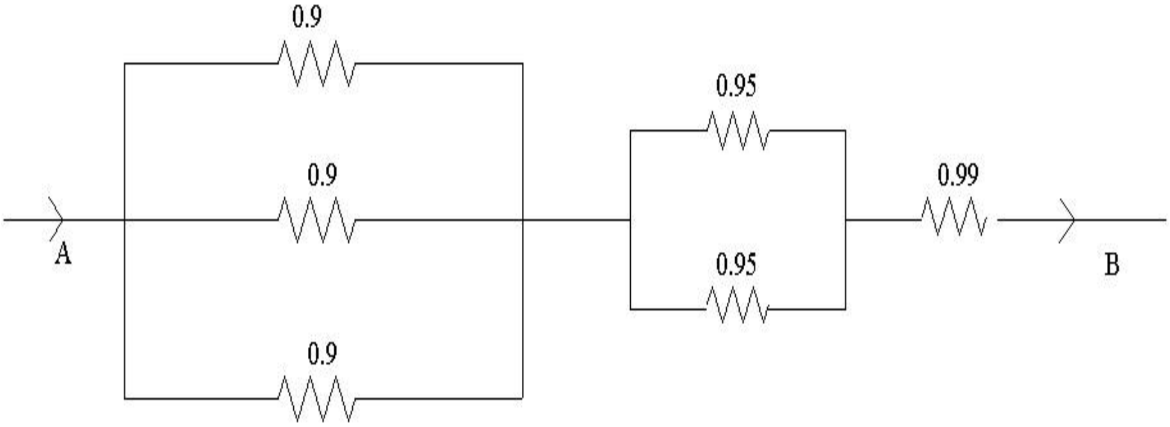
เขียน Tree Diagram ประกอบการคำนวณทั้งข้อ ก และ ข.

5. จงหาความน่าจะเป็นที่กระแสไฟฟ้าไหลผ่านวงจรไฟฟ้าดังรูป ถ้า A_i คือเหตุการณ์ที่หน่วยไฟฟ้าที่ i ทำงานได้ตามปกติ $P(A_i) = p$ และแต่ละหน่วยทำงานเป็นอิสระต่อกัน



เขียน Tree Diagram ประกอบ พร้อมเขียนเซทของเหตุการณ์ที่กระแสไฟฟ้าจะไหลผ่านวงจรได้

6. จงหาความน่าจะเป็นที่กระแสไฟฟ้าไหลผ่านวงจรไฟฟ้าจากจุด A ไป B ได้ ถ้ากำหนดความน่าจะเป็นที่แต่ละหน่วยของวงจรทำงานให้ดังแสดงในรูป และสมมติว่าแต่ละหน่วยวงจรทำงานอิสระต่อกัน



7. จากการสำรวจพบว่า 5 % ของประชากรในเมือง ๆ หนึ่ง เป็นผู้มีความดันโลหิตสูง และพบว่าในกลุ่มประชากรที่มีความดันโลหิตสูง 75 % เป็นผู้ดื่มสุรา และในกลุ่มประชากรที่ไม่ได้เป็นผู้มีความดันโลหิตสูง 50 % เป็นผู้ดื่มสุรา ถ้าพบชายผู้หนึ่งเป็นผู้มีความดันโลหิตสูงจงหาความน่าจะเป็นที่ชายผู้นี้จะเป็นผู้ดื่มสุรา

8. การแจกแจงความน่าจะเป็นของสัปดาห์ที่การนัดหยุดงานสามารถตกลงกันได้ ของโรงงานอุตสาหกรรมประเภทหนึ่งแสดงดังตาราง

สัปดาห์ที่สามารถตกลงกันได้	1	2	3	4	5
ความน่าจะเป็น	0.63	0.23	0.09	0.04	0.02

(ก) ถ้าการนัดหยุดงานเพิ่งเริ่มต้น (สัปดาห์ที่ 1) จงหาความน่าจะเป็นที่จะสามารถตกลงกันได้ในสัปดาห์ 1 ? สามารถตกลงกันได้ในสัปดาห์ที่ 5

(ข) ถ้าการนัดหยุดงานเข้าสู่สัปดาห์ที่ 2 จงหาความน่าจะเป็นที่จะสามารถตกลงกันได้ในสัปดาห์นี้

(ค) จงคำนวณหาความน่าจะเป็นที่การนัดหยุดงานตกลงกันไม่ได้ในสัปดาห์ที่ 1 และตกลงกันได้ในสัปดาห์ที่ 2

(ง) ถ้าการนัดหยุดงานเริ่มเข้าสู่สัปดาห์ที่ 3, 4 และ 5 จงหาความน่าจะเป็นที่การนัดหยุดงานจะตกลงกันได้ในสัปดาห์นั้น ถ้ากำหนดว่าไม่สามารถตกลงกันได้ในสัปดาห์ก่อนหน้านั้น

(จ) ถ้าเวลา (สัปดาห์) ผ่านไป การตกลงจะง่ายหรือยากขึ้น

9. บริษัทผู้ผลิตคอมพิวเตอร์ทำการตรวจสอบ Memory Chip ก่อนการประกอบโดยกำหนดให้ D คือชิพที่ไม่ได้มาตราฐาน และ N คือชิพที่ได้มาตราฐาน และให้ A คือเหตุการณ์ที่ชิพผ่านการตรวจสอบโดยวิศวกรบริษัท และ A' คือเหตุการณ์ที่ซิพไม่ผ่านการตรวจสอบ จากข้อมูลของบริษัทพบว่า

$$P(D) = 0.1$$

$$P(A|D) = 0.005$$

$$P(A|N) = 0.999$$

จงเขียน Tree Diagram อธิบายความน่าจะเป็นของเหตุการณ์ต่าง ๆ และจงหาว่า

(ก) ความน่าจะเป็นที่ซิพจะผ่านการตรวจสอบ

(ข) ความน่าจะเป็นที่ซิพไม่มาตราฐาน ถ้าซิพนั้นผ่านกาตรวจสอบ

10. คน 9 คนกำลังจะไปเที่ยวชายทะเลโดยใช้รถยนต์ 3 คัน ซึ่งสามารถจุคนได้ 2, 4 และ 5 คน ตามลำดับ จงหาว่าสามารถจัดคน 9 คนไปเที่ยวชายทะเลโดยใช้รถยนต์ทั้ง 3 คันได้กี่วิธี

บทที่ 2

ตัวแปรสุ่มและการคาดคะเนทางคณิตศาสตร์

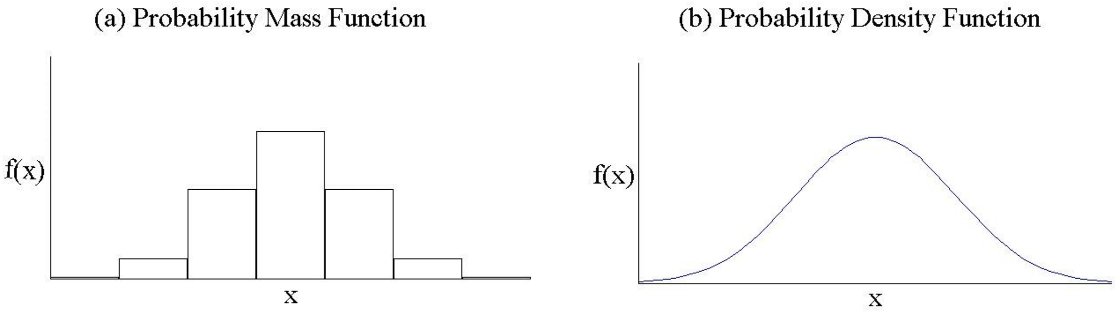
(Random Variable and Mathematical Expectation)

2.1 คำนำ

ตัวแปรสุ่ม คือ ตัวแปรที่แสดงผลลัพธ์การทดลอง ***ที่เราต้องการทราบ เป็นเลขจำนวนจริง*** ไม่สามารถบอกค่าล่วงหน้าได้ แต่สามารถอธิบายค่าของตัวแปรสุ่มได้ด้วยฟังก์ชั่นการแจกแจงความน่า จะเป็น ปกติเขียนแทนด้วยตัวใหญ่

ให้ X = ตัวแปรสุ่ม

f(x) = ฟังก์ชั่นการแจกแจงความน่าจะเป็น (Probability Distribution Function) ของตัวแปรสุ่ม X ดังแสดงในรูปที่ 2.1



รูปที่ 2.1 ฟังก์ชั่นการแจกแจงความน่าจะเป็น f(x)

เวลาพูดถึงตัวแปรสุ่ม X จะต้องรู้ว่าค่าของตัวแปรสุ่ม X มีค่าเป็นอะไรได้บ้าง และ f(x) มีรูปร่างอย่างไร

ค่าของตัวแปรสุ่มอาจเหมือนกับเอกภพสัมพัทธ์ (Sample Space) หรือเป็นค่าที่คำนวณจากเอกภพสัมพัทธ์ก็ได้ เช่น

การทดลอง 1: การทอดลูกเต๋า 1 ลูก

S = {1, 2, 3, 4, 5, 6}

ให้ X = ตัวแปรสุ่ม แสดงผลของการทอดลูกเต๋า 1 ลูก

x = 1, 2, 3, 4, 5, 6 เหมือน Sample Space

f(x) = $\frac{1}{6}$ เมื่อ x = 1, 2, ..., 6

การทดลอง 2 : การทอดลูกเต๋า 2 ลูกพร้อมกัน

$$S = \begin{Bmatrix} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{Bmatrix}$$

ให้ X = ผลบวกของการทอดลูกเต๋า 2 ลูกพร้อมกัน ซึ่งเป็นสิ่งที่เราต้องการทราบ

X	2	3	4	5	6	7	8	9	10	11	12
f(x)	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

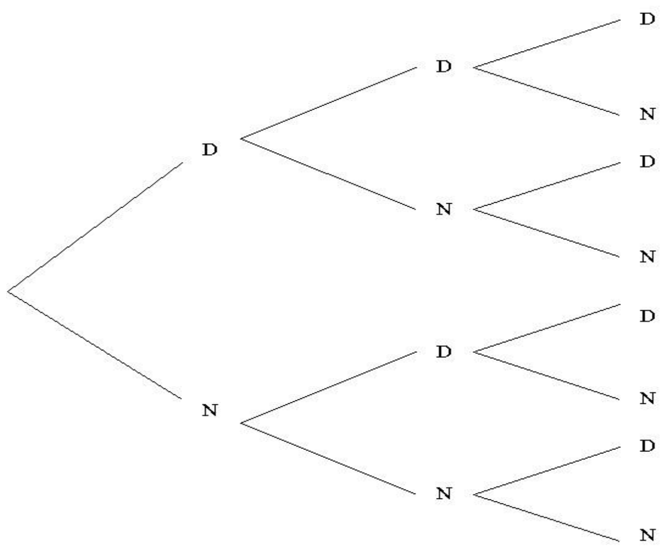
$$f(x) = \frac{6 - |x - 7|}{36} \quad \text{เมื่อ} \quad x = 2, 3, ..., 12$$

กรณีนี้ X จะคำนวณจากเอกภพสัมพัทธ์ของการทอดลูกเต๋า 2 ลูกพร้อมกัน

การทดลอง 3 : สุ่มตัวอย่างผลิตภัณฑ์ 3 ชิ้น การสุ่มแต่ละครั้งจะได้ผลคือ D (ผลิตภัณฑ์ชำรุด, Defective) หรือ N (ผลิตภัณฑ์ไม่ชำรุด, Non-defective) การสุ่มแต่ละครั้งเป็นอิสระต่อกัน หรือ P(D)และ P(N) มีค่าคงที่

ให้ P(D) = p

P(N) = 1-p



Sample Point	**Probability**
DDD	p^3
DDN	$p^2(1-p)$
DND	$p^2(1-p)$
DNN	$p(1-p)^2$
NDD	$p^2(1-p)$
NDN	$p(1-p)^2$
NND	$p(1-p)^2$
NNN	$(1-p)^3$
P(S) =	1.00

ถ้าเราสนใจจำนวนผลิตภัณฑ์ชำรุด (D) ที่สุ่มได้

X = ตัวแปรสุ่มแสดงจำนวนผลิตภัณฑ์ชำรุดที่สุ่มได้จากการสุ่มผลิตภัณฑ์ 3 ชิ้น ซึ่งเป็นตัวแปรสุ่มแบบทวินาม (Binomial) ซึ่งจะกล่าวถึงรายละเอียดในบทที่ 4

n = 3 (จำนวนชิ้นผลิตภัณฑ์ที่สุ่ม)

x = 0, 1, 2, 3

f(x) = $\binom{n}{x} p^{x} (1-p)^{n-x}$

x	0	1	2	3
F(x)	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

ตัวแปรสุ่มมี 2 ประเภทคือ

(1) ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)

(2) ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

2.2 ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)

ตัวแปรสุ่มไม่ต่อเนื่อง คือตัวแปรสุ่มที่มีค่าเป็นจำนวนนับได้ เช่น ผลบวกของการทอดลูกเต๋า 2 ลูกพร้อมกัน จำนวนผลิตภัณฑ์ชำรุดที่สุ่มได้จากการสุ่มผลิตภัณฑ์ 3 ชิ้น จำนวนคอมพิวเตอร์ที่ผลิตได้ในแต่ละเดือน เป็นต้น

คุณสมบัติสำคัญที่ควรทราบเกี่ยวกับตัวแปรสุ่มไม่ต่อเนื่อง

(1) $f(x) \geq 0$ for all x

$f(x)$ คือ Probability Mass Function (pmf) หรือเรียกสั้น ๆ ว่า Probability Function

(2) $\sum_{x} f(x) = 1$

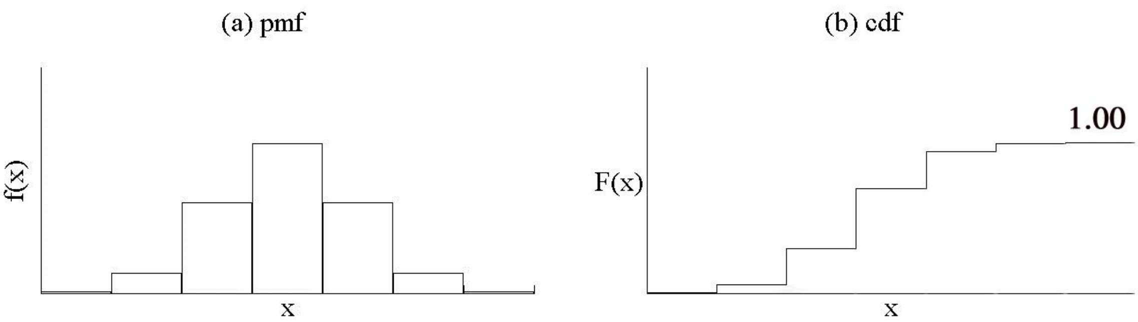
(3) $P(X=x) = f(x)$

(4) $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$ = Cumulative Probability Distribution Function หรือเรียกสั้น ๆ ว่า Cumulative Distribution Function (cdf) ดังแสดงในรูปที่ 2.2

= Non-exceedence Probability

(5) $P(X>x) = 1-P(X \leq x)$ = Exceedence Probability

(6) $P(a < X \leq b) = F(b)-F(a)$



รูปที่ 2.2 กราฟแสดง pmf และ cdf

ตัวอย่างที่ 2.1 หลอดไฟ 10 ดวง เสีย 3 ดวง ถ้าสุ่มเลือกหลอดไฟ 3 ดวง จงหาการแจกแจงความน่าจะเป็นของหลอดไฟที่เสีย

ให้ X = ตัวแปรสุ่ม แสดงจำนวนหลอดไฟที่เสียจากการสุ่มเลือกหลอดไฟ 3 ดวง (n = 3) x = 0, 1, 2, 3

k = จำนวนหลอดไฟเสีย = 3

$$f(x) = P(X = x) = \frac{\binom{3}{x}\binom{7}{n-x}}{\binom{10}{n}} \quad \text{เมื่อ } x = 0, 1, \ldots\ldots, \text{Min}(n,k)$$

x	0	1	2	3
f(x)	$35/120$	$63/120$	$21/120$	$1/120$

* กรณีนี้ไม่ใช้ Binomial Trial เนื่องจากความน่าจะเป็นที่จะได้หลอดเสียในการสุ่มแต่ละครั้งต่างกัน (Not independent) เช่น P(หลอดเสียในการหยิบครั้งแรก) = $\frac{3}{10}$, P(หลอดเสียในการหยิบครั้งที่ 2) = $\frac{3}{9}$ หรือ $\frac{2}{9}$ ขึ้นอยู่กับผลการหยิบครั้งแรกได้หลอดที่ดีหรือเสีย เป็นต้น

ตัวอย่างที่ 2.2 ในการขนส่งคอมพิวเตอร์ 8 เครื่อง โดยทางเรือปรากฏว่ามี 3 เครื่องชำรุด ถ้าโรงเรียนแห่งหนึ่งสั่งซื้อ 2 เครื่อง โดยสุ่ม จงหาการแจกแจงความน่าจะเป็นของจำนวนเครื่องคอมพิวเตอร์ชำรุดที่ทางโรงเรียนจะได้รับ

ให้ X = ตัวแปรสุ่มแสดงจำนวนเครื่องคอมพิวเตอร์ชำรุดที่ทางโรงเรียนจะได้รับ เมื่อสั่งซื้อ 2 เครื่อง (n=2)

x = 0, 1, 2

k = จำนวนคอมพิวเตอร์ชำรุด = 3

$$f(x) = P(X=x) = \frac{\binom{3}{x}\binom{5}{n-x}}{\binom{8}{n}} \quad \text{เมื่อ } x = 0, 1, \ldots\ldots\ldots\ldots, \text{Min}(k,n)$$

x	f(x)
0	$\dfrac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \dfrac{10}{28}$
1	$\dfrac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \dfrac{15}{28}$
2	$\dfrac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \dfrac{3}{28}$
	1.00

* ในตัวอย่างที่ 2.1 และ 2.2 ถ้าจำนวนหลอดไฟทั้งหมด หรือจำนวนคอมพิวเตอร์ทั้งหมด (N) มีค่ามาก เมื่อเทียบกับค่า n P(probability of success) จะมีค่าคงเกือบที่ในการสุ่มแต่ละครั้ง และสามารถใช้ทฤษฎี Binomial คำนวณหาการแจกแจงความน่าจะเป็นของ X ได้

ตัวอย่างที่ 2.3 ถ้าร้อยละ 50 ของรถยนต์ที่จำหน่ายโดยตัวแทนจำหน่ายเป็นเครื่องยนต์ดีเซล ถ้าตัวแทนจำหน่ายรถ 4 คัน จงหาการแจกแจงความน่าจะเป็นสะสมของรถยนต์ดีเซลที่จำหน่ายออกไป

ให้ X = ตัวแปรสุ่มแสดงจำนวนรถยนต์ดีเซลที่จำหน่ายจากรถทั้งหมด 4 คัน เป็นตัวแปรสุ่มแบบ Binomial ซึ่งมี n=4

x = 0, 1, 2, 3, 4

p = $\dfrac{1}{2}$

f(x) = $\binom{n}{x} p^x (1-p)^{n-x}$ เมื่อ x = 0, 1, ..., 4

x	f(x)	F(x)
0	$\binom{4}{0}\left(\frac{1}{2}\right)^4 = \frac{1}{16}$	$\frac{1}{16}$
1	$\binom{4}{1}\left(\frac{1}{2}\right)^4 = \frac{4}{16}$	$\frac{5}{16}$
2	$\binom{4}{2}\left(\frac{1}{2}\right)^4 = \frac{6}{16}$	$\frac{11}{16}$
3	$\binom{4}{3}\left(\frac{1}{2}\right)^4 = \frac{4}{16}$	$\frac{15}{16}$
4	$\binom{4}{4}\left(\frac{1}{2}\right)^4 = \frac{1}{16}$	$\frac{1}{16}$
		1.00

2.3 ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

ตัวแปรสุ่มต่อเนื่อง คือ ตัวแปรสุ่มซึ่งมีค่าต่อเนื่อง จนไม่สามารถนับจำนวนค่าของตัวแปรที่อยู่ระหว่าง 2 ค่าใด ๆ ได้ เช่นค่าความสูง น้ำหนัก อุณหภูมิ ระยะทาง หรืออายุการใช้งานของอุปกรณ์ เป็นต้น ความน่าจะเป็นของการเกิดค่าใดค่าหนึ่งของตัวแปรสุ่มต่อเนื่องจะมีค่าเท่ากับ 0 หรือ $P(X=x) = 0$

คุณสมบัติสำคัญที่ควรทราบเกี่ยวกับตัวแปรสุ่มต่อเนื่องได้แก่

1. $f(x) \geq 0$ for all x (Probability Density Function, pdf)
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx =$ Cumulative Probability Distribution Function (cdf)
4. $P(a < X \leq b) = \int_{a}^{b} f(x)\, dx = F(b)-F(a)$

ตัวอย่างที่ 2.4 ค่าความผิดพลาดของอุณหภูมิ ($^{\circ}$C) ในกระบวนการทางเคมีเป็นตัวแปรสุ่มต่อเนื่อง มี pdf ดังนี้

$$f(x) = \frac{x^2}{3} \quad \text{เมื่อ } -1 \leq x < 2$$

(1) จงทดสอบว่า f(x) เป็น pdf จริง

(2) จงหา $P(0 \leq X < 1)$

(3) จงหา F(x)

(4) จงหา $P(0 \leq X < 1)$ จาก F(x)

(1) $f(x) = \frac{x^2}{3} \geq 0$ สำหรับทุกค่าของ x

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{2} \frac{x^2}{3}\,dx$$
$$= \frac{x^3}{9} \Big|_{-1}^{2}$$
$$= \frac{1}{9}(8-(-1)) = 1$$

แสดงว่า $f(x) = \frac{x^2}{3}$ เป็น pdf ของ x จริง

(2) $P(0 \le X < 1) = \int_{0}^{1} f(x)\,dx$
$$= \int_{0}^{1} \frac{x^2}{3}\,dx$$
$$= \frac{x^3}{9} \Big|_{0}^{1} = \frac{1}{9}(1-0) = \frac{1}{9}$$

(3) $F(x) = \int_{-\infty}^{x} f(x)\,dx$
$$= \int_{-1}^{x} \frac{x^2}{3}\,dx$$
$$= \frac{x^3}{9} \Big|_{-1}^{x}$$
$$= \frac{1}{9}(x^3+1)$$

(4) $P(0 \le X < 1) = F(1) - F(0)$
$$= \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

2.4 การแจกแจงความน่าจะเป็นของผลการทดลอง (Empirical Distributions)

ในหัวข้อที่แล้วได้กล่าวถึง Probability Mass Function และ Probability Density Function, f(x) ซึ่งเป็นคุณสมบัติการแจกแจงความน่าจะเป็นของประชากร (Population) ซึ่งปกติจะไม่ทราบ

การทดลองทางสถิติจะทำให้ได้ข้อมูลเกี่ยวกับตัวแปรสุ่ม ซึ่งสามารถนำมาวิเคราะห์การแจกแจงความถี่ได้โดยวิธีใดวิธีหนึ่งดังต่อไปนี้

- Stem and Leaf Plot
- Double Stem and Leaf Plot
- ตารางการแจกแจงความถี่สัมพัทธ์
- กราฟการแจกแจงความถี่สัมพัทธ์ (Relative Frequency Histogram)
- ตารางการแจกแจงความถี่สัมพัทธ์สะสม
- กราฟการแจกแจงความถี่สัมพัทธ์สะสม

โดยการลาก Smooth Curve ผ่าน Relative Frequency Histogram จะได้รูปร่างโดยประมาณของ Probability Density Function, f(x) สามารถใช้ความรู้ทางคณิตศาสตร์เพื่อหาสมการ

ของ f(x) ได้ซึ่งปกติสมการจะอยู่ในรูปของสมการ Parabolas, Hyperbolas, Hypergeometric หรือ Exponential และสุดท้ายคือการหาพารามิเตอร์ของสมการจากข้อมูลที่ได้จากการทดลอง

ตัวอย่างที่ 2.5 ในการศึกษาอายุการใช้งานของแบตเตอรี่รถยนต์ ได้ทำการสุ่มตัวอย่างอายุการใช้งานของรถยนต์ จำนวน 40 ตัวอย่าง ดังตาราง

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.3	2.6	3.9	3.0	4.2	3.5

จัดทำตาราง Stem and Leaf Plot ได้ดังนี้

จำนวนเต็ม (Stem)	ทศนิยม (Leaves)	ความถี่ (Frequency)
1	69	2
2	25696	5
3	4318514723628297130097145	25
4	71354172	8

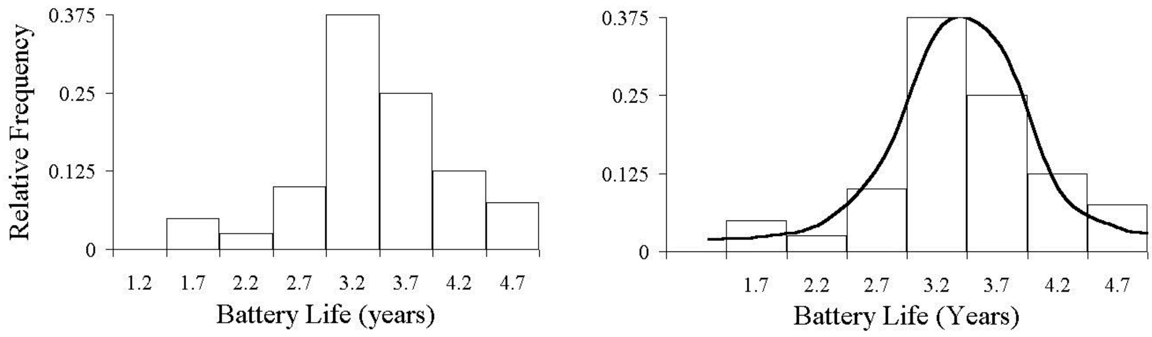
จากตารางจะเห็นได้ว่า Stem and Leaf Plot ข้างบนมีข้อมูลไม่เพียงพอที่จะเห็นภาพของการแจกแจงชัดเจน จึงต้องเพิ่มจำนวน Stem ให้มากขึ้น โดยแบ่ง Leaf ออกเป็น 2 พวก คือ 0, 1, 2, 3, 4 และ 5, 6, 7, 8, 9 และสร้างตาราง Double-Stem and Leaf Plot ได้ดังนี้

จำนวนเต็ม (Stem)	ทศนิยม (Leaves)	ความถี่ (Frequency)
1	69	2
2	2	1
2	5696	4
3	431142322130014	15
3	8576897975	10
4	13412	5
4	757	3

จากตาราง Double-Stem and Leaf Plot จะเห็นการแจกแจงความถี่ได้ชัดเจนขึ้น และสามารถสร้างเป็นตารางการแจกแจงความถี่สัมพัทธ์ได้ดังนี้

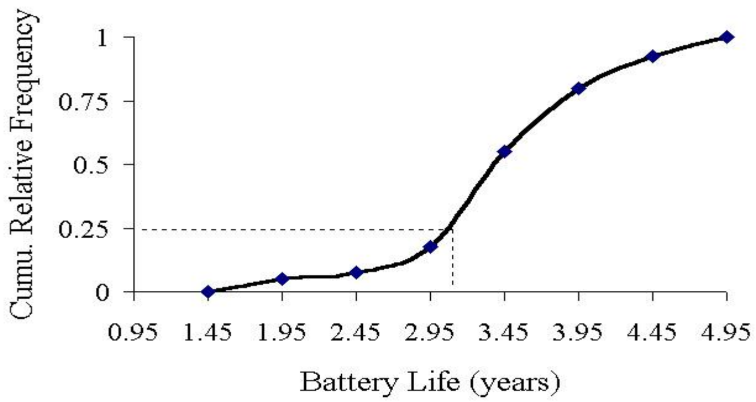
อันตรภาคชั้น (Class Interval)	จุดกึ่งกลางชั้น (Class Midpoint)	ความถี่ (Frequency)	ความถี่สัมพัทธ์ (Relative frequency)
1.5 - 1.9	1.7	2	0.050
2.0 - 2.4	2.2	1	0.025
2.5 - 2.9	2.7	4	0.100
3.0 - 3.4	3.2	15	0.375
3.5 - 3.9	3.7	10	0.250
4.0 - 4.4	4.2	5	0.125
4.5 - 4.9	4.7	3	0.075

จากตารางการแจกแจงความถี่สัมพัทธ์ สามารถสร้างกราฟการแจกแจงความถี่สัมพัทธ์ได้ดังรูปที่ 2.3



รูปที่ 2.3 กราฟการแจกแจงความถี่สัมพัทธ์ และ f(x) ของอายุการใช้งานของแบตเตอรี่

ทำการ Smooth Curve จะได้ f(x) และสามารถสร้างกราฟการแจกแจงความถี่สัมพัทธ์สะสมได้ดังนี้



ความรู้เกี่ยวกับการแจกแจงความน่าจะเป็นของตัวแปรสุ่มจะช่วยตอบคำถาม 2 ข้อ คือ

(1) ช่วยบอกความน่าจะเป็นของเหตุการณ์ที่เราต้องการทราบ เช่น

P (อายุการใช้งานของแบตเตอรี่ยี่ห้อหนึ่ง > 2 ปี) = x

ค่า x จะมีประโยชน์ต่อผู้ใช้ในการเลือกซื้อแบตเตอรี่

(2) ช่วยบอกค่าของเหตุการณ์ที่ความน่าจะเป็นที่กำหนด เช่น

P (อายุการใช้งานของแบตเตอรี่ยี่ห้อหนึ่ง > y ปี) = 0.9

ค่า y จะมีประโยชน์ต่อผู้ผลิตแบตเตอรี่ในการกำหนดอายุประกันการใช้งาน (Warrantee Period) ของแบตเตอรี่

2.5 การแจกแจงความน่าจะเป็นร่วมกัน (Joint Probability Distribution)

ถ้าผลการทดลองขึ้นอยู่กับตัวแปรสุ่มสองตัว (X, Y) f(x, y) = Joint Probability Distribution ของ X และ Y ซึ่งก็คือ P(X=x, Y=y) หรือ $P(X \cap Y)$ สำหรับกรณีที่ X, Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง

2.5.1 Discrete Joint Probability Distributions

คุณสมบัติที่สำคัญที่ควรทราบคือ

(1) $f(x, y) \geq 0$ สำหรับทุกค่าของ x, y

(2) $\sum_x \sum_y f(x, y) = 1$

(3) $P(X = x, Y = y) = f(x, y)$

(4) $P[(X, Y) \subset A] = \sum\sum_A f(x, y)$

ตัวอย่าง 2.6 ถุงใบหนึ่งบรรจุผลไม้ 3 ชนิด คือ ส้ม 3 ผล มะม่วง 2 ผล และมังคุด 3 ผล สุ่มเลือกผลไม้ 4 ผล ถ้าให้ X แทนจำนวนส้ม และ Y แทนจำนวนมะม่วงที่หยิบได้จงหา

ก. การแจกแจงความน่าจะเป็นร่วมกันของ X และ Y

ข. $P[(X, Y) \subset A]$ เมื่อ $A = [(x,y) \mid x + y \leq 2]$

ค. $\sum_x \sum_y f(x,y)$

$$P(X=x, Y=y) = f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{4-x-y}}{\binom{8}{4}}$$

เมื่อ $x = 0, 1, 2, 3$

$y = 0, 1, 2$

และ $0 \leq x + y \leq 4$

ก. ตาราง f(x,y)

y \ x	0	1	2	3	รวม
0	-	3/70	9/70	3/70	15/70
1	2/70	18/70	18/70	2/70	40/70
2	3/70	9/70	3/70	-	15/70
รวม	5/70	30/70	30/70	5/70	1

ข. $P[(X,Y \subset A]$ เมื่อ $A = [(x,y) \mid x + y \le 2]$

$$= f(0,0) + f(0,1) + f(0,2) + f(1,0) + f(1,1) + f(2,0)$$

$$= 0 + \frac{2}{70}+\frac{3}{70}+\frac{3}{70}+\frac{18}{70}+\frac{9}{70}$$

$$= \frac{1}{2}$$

ค. $$\sum_x \sum_y f(x,y) = \frac{3}{70}+\frac{9}{30}+\frac{3}{70}+\frac{2}{70}+\frac{18}{70}+\frac{18}{70}+\frac{2}{70}+\frac{3}{70}+\frac{9}{70}+\frac{3}{70}$$

$$= 1$$

2.5.2 Continuous Joint Probability Distributions

คุณสมบัติสำคัญที่ควรทราบคือ

(1) $f(x, y) \geq 0$ สำหรับทุกค่าของ x,y

(f(x,y) = Joint Probability Density Function)

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$

(3) $P[(X, Y) \subset A] = \iint_A f(x, y)\, dx\, dy$

ตัวอย่าง 2.7 กำหนด Joint Probability Density Function ของตัวแปรสุ่มต่อเนื่อง X,Y คือ

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & \text{เมื่อ } 0 \le x < 1; 0 \le y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่นๆ} \end{cases}$$

ก. จงทดสอบว่า f(x,y) เป็นฟังก์ชันความน่าจะเป็นร่วมกันของตัวแปรสุ่มต่อเนื่อง X, Y จริง

ข. จงหา $P(0 \le X < 1/2, 1/4 \le Y < 1/2)$

ก. $$f(x,y) = \begin{cases} \frac{2}{5}(2x + 3y) & \text{เมื่อ } 0 \le x < 1 \;;\; 0 \le y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่นๆ} \end{cases}$$

ทุกค่า x, y ทำให้ $f(x,y) \ge 0$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = \int_0^1 \int_0^1 \frac{2}{5}(2x + 3y).\, dx.\, dy$$

$$= \int_0^1 \frac{2x^2}{5}+\frac{6xy}{5} \Big|_0^1 \, dy$$

$$= \int_0^1 (\frac{2}{5}+\frac{6y}{5}).\, dy$$

$$= \frac{2y}{5}+\frac{3y^2}{5}\Big|_0^1$$

$$= 1$$

f(x,y) เป็นฟังก์ชันความน่าจะเป็นร่วมกันของตัวแปรสุ่มต่อเนื่อง X,Y

ข. P(0 < X < 1/2, 1/4 < Y < 1/2)

$$= \int_{\frac{1}{4}}^{\frac{1}{2}}\int_{0}^{\frac{1}{2}} \frac{2}{5}(2x + 3y).\ dx.\ dy$$

$$= \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{2x^2}{5} + \frac{6xy}{5}\Big|_{0}^{\frac{1}{2}}\ dy$$

$$= \int_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{1}{10} + \frac{3y}{5}\right) dy$$

$$= \frac{y}{10} + \frac{3y^2}{10}\Big|_{\frac{1}{4}}^{\frac{1}{2}}$$

$$= \frac{13}{160}$$

2.6 Marginal Probability Distributions

Marginal Probability Distributions คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มตัวใดตัวหนึ่ง เช่น X หรือ Y ซึ่งคำนวณจาก Joint Probability Distributions

คุณสมบัติของ Marginal Probability Distributions ที่ควรทราบคือ

Marginal Distribution of X	**Marginal Distribution of Y**
(1) $g(x) = \sum_{y} f(x, y)$ ....(Discrete) $= \int_{y} f(x, y)\, dy$...(Continuous)	$h(y) = \sum_{y} f(x, y)$.......(Discrete) $= \int_{x} f(x, y)\, dx$ ...(Continuous)
(2) $\sum_{x}\sum_{y} f(x, y) = \sum_{x} g(x) = 1$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy$ $= \int_{-\infty}^{\infty} g(x)\, dx = 1$	$\sum_{x}\sum_{y} f(x, y) = \sum_{y} h(y) = 1$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy$ $= \int_{-\infty}^{\infty} h(y)\, dy = 1$
(3) $f(x, y) = g(x)\, h(y)$	ถ้า X และ Y (Independent)
(4) $f(x, y) = g(x)\, f(y\|x)$ $= h(y)\, f(x\|y)$	ถ้า Y ขึ้นกับ X (Conditional) ถ้า X ขึ้นกับ Y (Conditional)

2.7 การแจกแจงแบบมีเงื่อนไข (Conditional Probability Distributions)

กรณีที่ตัวแปรสุ่ม X และ Y ไม่เป็นอิสระต่อกัน การแจกแจงความน่าจะเป็นของตัวแปรสุ่มตัวหนึ่งเช่น Y จะขึ้นอยู่กับค่าของตัวแปรสุ่ม X ว่ามีค่าเท่ากับเท่าใด เช่น

$$P(Y = y|\, X = x) = \frac{P(X = x,\, Y = y)}{P(X = x)}$$

คุณสมบัติที่สำคัญของ Conditional Probability Distributions ได้แก่

$$f(y|x) = \frac{f(x, y)}{g(x)} \quad \text{เมื่อ } g(x) > 0$$

$$f(x|y) = \frac{f(x, y)}{h(y)} \quad \text{เมื่อ } h(y) > 0$$

ตัวอย่าง 2.8 กำหนดให้ X และ Y ตัวแปรสุ่มไม่ต่อเนื่อง มีการแจกแจงความน่าจะเป็นร่วมกันดังแสดงในตาราง จงหา

ก. Marginal Probability Distributions ของ X และ Y

ข. จงหา P(Y= 1|X = 2)

y \ x	1	2	3
1	0	1/6	1/12
2	1/12	1/9	0
3	1/4	1/4	1/18

ก. Marginal Distribution Function g(x) และ h(y)

y \ x	1	2	3	h(y)
1	0	1/6	1/12	9/36
2	1/12	1/9	0	7/36
3	1/4	1/4	1/18	20/36
g(x)	12/36	19/36	5/36	1.00

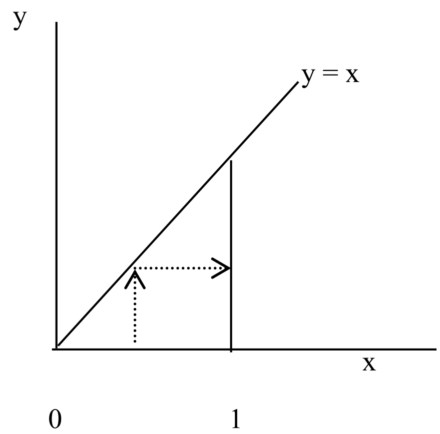
ข. $P(Y = 1|X = 2] \quad = \quad \dfrac{f(2,1)}{g(2)}$

$$= \frac{\frac{1}{6}}{\frac{19}{36}}$$

$$= \frac{36}{6x19} = \frac{6}{19}$$

ตัวอย่าง 2.9 ถ้า X, Y เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมีฟังก์ชันการแจกแจงความน่าจะเป็นร่วมกันดังนี้คือ

$$f(x,y) = \begin{cases} 8xy & \text{เมื่อ } 0 \le x < 1 \; ; \; 0 \le y < x \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่นๆ} \end{cases}$$



จงหา g(x), h(y), f(y|x), $P(Y < \frac{1}{8} \,|X = \frac{1}{2})$

$$g(x) = \int_0^x 8xy.\, dy$$

$$g(x) = 4xy^2 \Big|_0^x$$

$$g(x) = \begin{cases} 4x^3 & \text{เมื่อ } 0 \le x < 1 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่นๆ} \end{cases}$$

$$h(y) = \int_y^1 8xy.\, dx$$

$$h(y) = 4x^2y \Big|_y^1$$

$$h(y) = \begin{cases} 4y(1 - y^2) & \text{เมื่อ } 0 \le y < 1 \\ 0 & \text{เมื่อ } y \text{ มีค่าอื่นๆ} \end{cases}$$

$$f(y|x) = \frac{f(x,y)}{g(x)}$$

$$f(y|x) = \frac{8xy}{4x^3}$$

$$f(y|x) = \begin{cases} \frac{2y}{x^2} & \text{เมื่อ } 0 \le y < x \\ 0 & \text{เมื่อ } y \text{ มีค่าอื่น ๆ} \end{cases}$$

$$P(Y < \frac{1}{8} | X = \frac{1}{2}) = \int_0^{\frac{1}{8}} \frac{2y}{(\frac{1}{2})^2} . dy$$

$$= \int_0^{\frac{1}{8}} 8y. dy$$

$$= \frac{1}{16}$$

2.8 การเป็นอิสระของตัวแปรสุ่ม (Independence)

X เป็นอิสระ (Independent) จาก Y ถ้า

$$f(x|y) = g(x)$$

หรือ Y เป็นอิสระจาก X ถ้า

$$f(y|x) = h(y)$$

จาก $f(x|y) = \frac{f(x, y)}{h(y)}$

และ $f(y|x) = \frac{f(x, y)}{g(x)}$

จะได้ว่า $f(x, y) = g(x). h(y)$ ถ้า X และ Y เป็นอิสระต่อกัน

ตัวอย่าง 2.10 กำหนดให้ X, Y มีการแจกแจงความน่าจะเป็นร่วมกันดังแสดงในตาราง

y \ x	2	4
1	0.10	0.15
2	0.20	0.30
3	0.10	0.15

จงตรวจสอบว่า X, Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน

y \ x	2	4	h(y)
1	0.1	0.15	0.25
3	0.2	0.30	0.5
5	0.1	0.15	0.25
g(x)	0.4	0.6	1.00

$f(x,y) = g(x) \cdot h(y)$ ทุกค่าของ x,y

ดังนั้น X, Y เป็นตัวแปรสุ่มไม่ต่อเนื่องซึ่งเป็นอิสระต่อกัน

ตัวอย่าง 2.11 กำหนดให้ฟังก์ชันความน่าจะเป็นร่วมกันของตัวแปรสุ่ม X และ Y

$$f(x,y) = \begin{cases} 8xy & \text{เมื่อ } 0 \leq y < x \quad ; \quad 0 \leq x < 1 \\ 0 & \end{cases}$$

จงตรวจสอบว่า X, Y เป็นอิสระต่อกัน

$$\begin{aligned}
g(x) &= \int_0^x f(x, y)\, dy \\
&= \int_0^x 8xy \,.\, dy \\
&= 4x^3 \\
h(y) &= \int_y^1 8xy \,.\, dx \\
&= 4y(1-y^2) \\
g(x) \cdot h(y) &= 4x^3 [4y(1-y^2)] \\
g(x) \cdot h(y) &= 16x^3 y(1-y^2) \\
f(x, y) &\neq g(x) \cdot h(y)
\end{aligned}$$

ดังนั้นตัวแปรสุ่มต่อเนื่อง X, Y ไม่เป็นอิสระต่อกัน

2.9 กรณีตัวแปรสุ่มมากกว่า 2 ตัว

ถ้า $X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มซึ่งมีความน่าจะเป็นร่วมกัน (Joint Probability Distribution Function) เท่ากับ $f(x_1, x_2, \ldots, x_n)$

$f(x_1, x_2, \ldots\ldots\ldots x_n) = f(x_1).\ f(x_2)\ldots\ldots\ldots f(x_n)$ ถ้า $X_1, X_2, \ldots\ldots\ldots X_n$ เป็นอิสระต่อกัน

ตัวอย่าง 2.12 กำหนดให้ฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X, Y, Z มีค่าดังนี้

$$f(x, y, z) = \frac{4xyz^2}{9} \quad \text{เมื่อ } 0 \le x < 1,\ 0 \le y < 1 \text{ และ } 0 \le z < 3$$

$$= 0 \quad \text{เมื่อ } x, y, z \text{ มีค่าอื่น ๆ}$$

จงตรวจสอบว่าตัวแปรสุ่ม X, Y, Z เป็นอิสระต่อกัน

$$f_1(x) = \int_0^3 \int_0^1 \frac{4xyz^2}{9}\, dy\, .\, dz$$

$$= \frac{4x}{9} \int_0^3 \frac{y^2}{2} z^2 \Big|_0^1 dz$$

$$= \frac{2x}{9} \cdot \frac{z^3}{3} \Big|_0^3$$

$$= 2x \quad \text{เมื่อ } 0 \le x < 1$$

$$f_2(y) = \int_0^3 \int_0^1 \frac{4xyz^2}{9}\, dx\, .\, dz$$

$$= \frac{4y}{9} \int_0^3 \frac{x^2}{2} z^2 \Big|_0^1 dz$$

$$= 2y \quad \text{เมื่อ } 0 \le y < 1$$

$$f_3(z) = \int_0^1 \int_0^1 \frac{4xyz^2}{9}\, dx\, .\, dy$$

$$= \frac{4z^2}{9} \int_0^1 \frac{x^2 y}{2} \Big|_0^1 dy$$

$$= \frac{2z^2}{9} \int_0^1 y.\, dy$$

$$= \frac{z^2}{9} \quad \text{เมื่อ } 0 \le z < 1$$

$$f_1(x).\ f_2(y).\ f_3(z) = 2x.\ 2y.\ \frac{z^2}{9}$$

$$= \frac{4xyz^2}{9}$$

$$= f(x, y, z)$$

ดังนั้นตัวแปรสุ่ม X, Y, Z เป็นอิสระต่อกัน

2.10 การคาดคะเนทางคณิตศาสตร์ (Mathematical Expectation)

การคาดคะเนทางคณิตศาสตร์คือวิธีการคำนวณหาค่าพารามิเตอร์ของตัวแปรสุ่ม เช่น Mean, Variance และ Covariance โดยวิธีการคำนวณหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มหรือฟังค์ชั่นของตัวแปรสุ่ม

2.10.1 การหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มหรือฟังค์ชั่นของตัวแปรสุ่ม

ให้ X เป็นตัวแปรสุ่ม มีฟังค์ชั่นการแจกแจงความน่าจะเป็นเท่ากับ f(x)

ถ้า X เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$E(X) = \sum_{x} x.\ f(x) \quad \text{.......} \quad (2.1)$$

= Mean ของ X หรือ Mean ของฟังค์ชั่นการแจกแจงความน่าจะเป็นของ X

= μ_x = Moment ที่ 1 รอบจุดศูนย์

= เป็นพารามิเตอร์ที่แสดงจุดศูนย์กลางของการแจกแจง

ถ้า X เป็นตัวแปรสุ่มต่อเนื่อง

$$E(X) = \int_{x} x.\ f(x)\,dx \quad \text{.......} \quad (2.2)$$

ถ้า Z = u(X) = ฟังค์ชั่นใด ๆ ของ X

$$E(Z) = E(u(X)) = \mu_{u(X)}$$

$$= \sum_{x} u(x)\ f(x) \quad \text{....... (Discrete) ....} \quad (2.3)$$

$$= \int_{x} u(x).\ f(x)\,dx \quad \text{.......(Continuous).......} \quad (2.4)$$

ตัวอย่างที่ 2.13 ให้ X คือตัวแปรสุ่มแสดงผลของการทอดลูกเต๋า 1 ลูก จงหาค่าเฉลี่ยของ X

X	1	2	3	4	5	6	$\sum$
f(x)	1/6	1/6	1/6	1/6	1/6	1/6	1.00
x.f(x)	1/6	2/6	3/6	4/6	5/6	6/6	21/6

$$E(x) = \sum_{x=1}^{6} x.\ f(x)$$

$$= \frac{21}{6} = 3.5$$

ตัวอย่างที่ 2.14 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง มีฟังค์ชั่นความน่าจะเป็นดังนี้

$f(x) = 2(1-x)$ เมื่อ $0 \le x < 1$

จงหา E(X)

$$E(X) = \int_0^1 x\, f(x)\, dx$$
$$= \int_0^1 x.\ 2(1-x)\, dx$$
$$= 2\int_0^1 (x-x^2)\, dx$$
$$= 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]\Big|_0^1$$
$$= 2\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{3} = \mu_x$$

ตัวอย่างที่ 2.15 ในการขับรถแข่งครั้งหนึ่ง นักขับรถแข่งประกันรถในวงเงิน 200,000 บาท บริษัทประกันประมาณว่าความน่าจะเป็นที่จะต้องจ่ายค่าประกันระดับต่าง ๆ มีดังนี้

X = ค่าประกันที่บริษัทต้องจ่าย (บาท)

x (บาท)	200,000	100,000	50,000
f(x)	0.002	0.01	0.1

ถ้าบริษัทต้องการกำไร 2,000 บาท จะคิดค่าประกันเท่าใด

x (บาท)	200,000	100,000	50,000	0	$\sum$
f(x)	0.002	0.01	0.1	0.888	1.00
x. f(x)	400	1,000	5,000	0	6,400

$$E(X) = \sum_x x.\ f(x)$$
$$= 6{,}400 \text{ บาท } = \mu_x$$

$\therefore$ บริษัทต้องคิดค่าประกัน = 6,400 + 2,000 บาท

= 8,400 บาท

ตัวอย่างที่ 2.16 ให้ X คือจำนวนรถที่ผ่านเข้าไปในสถานีบริการในระหว่างเวลา 16.00-17.00 ของเย็นวันศุกร์ ด้วยความน่าจะเป็นดังนี้

X	4	5	6	7	8	9
f(x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

ถ้าให้ Z = จำนวนเงินที่รถคันหนึ่งต้องจ่ายเป็นค่าบริการ

= u(X) = 100 (2x-1)

จงหา E(Z)

x	4	5	6	7	8	9	$\sum$
f(x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$	1.00
x. f(x)	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{4}$	$\frac{7}{4}$	$\frac{8}{6}$	$\frac{9}{6}$	$\frac{82}{12}$
z	700	900	1,100	1,300	1,500	1,700	
z. f(x)	$\frac{700}{12}$	$\frac{900}{12}$	$\frac{1,100}{4}$	$\frac{1,300}{4}$	$\frac{1,500}{6}$	$\frac{1,700}{6}$	$\frac{15,200}{12}$

$$E(Z) = \sum_x z.\, f(x) = \frac{15,200}{12} = 1,267 \text{ บาท}$$

$$= \sum_x 100\,(2x-1)\, f(x)$$

$$= 100\,[2\sum_x .\, x\, f(x) - \sum_x f(x)]$$

$$= 100\,[2 \times \frac{82}{12} - 1] = 1,267 \text{ บาท}$$

2.10.2 กฎเกี่ยวกับ Expected Value

(1) $E(c) = c$ ………………………… (2.5)

(2) $E(c.\, u(X)) = c\, E(u(X))$ ………………………… (2.6)

(3) $E(\sum_{i=1}^{k} u_i(X)) = \sum_{i=1}^{k} E(u_i(X))$ ………………………… (2.7)

(4) $E((aX+b)^n) = \sum_{i=0}^{n} \binom{n}{i} a^{n-i}\, b^{i}\, E(X^{n-i})$ ……………… (2.8)

$E((aX+b)^2) = a^2\, E(X^2) + 2ab\, E(X) + b^2$

ตัวอย่างที่ 2.17 ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องซึ่งมีการแจกแจงความน่าจะเป็นดังตาราง

X	-3	6	9
f(x)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

จงหา $E(X)$, $E(X^2)$, $E(2X+1)^2$, $E[X-E(X)]^2$

X	-3	6	9	$\sum$
f(x)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	1.00
x. f(x)	$-\frac{1}{2}$	3	3	5.5
x^2	9	36	81	
x^2. f(x)	$\frac{9}{6}$	18	27	46.5

$$
\begin{aligned}
E(X) &= 5.5 \\
E(X^2) &= 46.5 \\
E(2X+1)^2 &= E(4X^2+4X+1) \\
&= 4E(X^2)+4E(X)+1 \\
&= 4 \times 46.5 + 4 \times 5.5 + 1 \\
&= 209 \\
E[X-E(X)]^2 &= E[X^2-2E(X).X+(E(X))^2] \\
&= E(X^2)-2E(X).E(X)+(E(X))^2) \\
&= E(X^2)-(E(X))^2 \\
&= 46.5-(5.5)^2 = 16.25 \quad \text{(Variance ของ X)}
\end{aligned}
$$

2.10.3 การหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มที่เป็นฟังค์ชั่นของตัวแปรสุ่มร่วมกัน 2 ตัว

ให้ X, Y เป็นตัวแปรสุ่ม มีฟังค์ชั่นความน่าจะเป็นร่วมกันเท่ากับ f(x,y)

ให้ Z = u(X,Y) = ฟังค์ชั่นใด ๆ ของ X, Y

$$E(Z) = E(u(X,Y)) = \mu_z = \mu_{u(X,Y)}$$

$$= \sum_y \sum_x u(x,y).\, f(x,y) \quad \text{.................. (Discrete)....... (2.9)}$$

$$= \int_y \int_x u(x,y).\, f(x,y)\, dxdy \quad \text{.................... (Continuous)… (2.10)}$$

$$E(X) = \mu_x = \sum_y \sum_x x\, f(x,y) \quad \text{.................... (Discrete)....... (2.11)}$$

$$= \sum_x x.\, g(x) \quad \text{.. (Discrete)....... (2.12)}$$

$$= \int_y \int_x x\, f(x,y)\, dx\, dy \quad \text{.................... (Continuous)... (2.13)}$$

$$= \int_x x.\, g(x)\, dx \quad \text{.................... (Continuous)... (2.14)}$$

$$E(Y) = \mu_y = \sum_y \sum_x y\, f(x,y) \quad \text{.................... (Discrete)....... (2.15)}$$

$$= \sum_y y.\, h(y) \quad \text{.............................. (Discrete)....... (2.16)}$$

$$= \int_y \int_x y\, f(x,y)\, dx\, dy \quad \text{.............................. (Continuous)... (2.17)}$$

$$= \int_y y.\, h(y)\, dy \quad \text{.............................. (Continuous)... (2.18)}$$

ถ้า $Z = XY$

$$E(Z) = E(XY)$$

$$= \sum_y \sum_x xy.\, f(x,y) \quad \text{.............................. (Discrete)...... (2.19)}$$

$$= \int_y \int_x xy.\, f(x,y)\, dx\, dy \quad \text{.............................. (Continuous)... (2.20)}$$

ถ้า X, Y เป็นอิสระต่อกัน (Independent)

$$f(x,y) = g(x).\, h(y)$$

$$E(XY) = E(X).\, E(Y) \quad \text{... (2.21)}$$

2.10.4 การหาค่าเฉลี่ยของตัวแปรสุ่มซึ่งเป็นฟังก์ชั่นของตัวแปรสุ่มร่วมกัน n ตัว

ให้ $X_1, X_2, \ldots\ldots, X_n$ เป็นตัวแปรสุ่ม n ตัว

ซึ่งมีฟังก์ชั่นความน่าจะเป็นร่วมกันคือ $f(x_1, x_2, \ldots\ldots, x_n)$

ให้ $Z = u(X_1, X_2, \ldots\ldots, X_n)$

2.10.4.1 กรณีที่ $X_1, X_2, \ldots\ldots, X_n$ เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$E(Z) = \sum_{x1} \sum_{x2} \ldots \sum_{xn} u(x_1,x_2,\ldots,x_n)\, f(x_1,x_2,\ldots,x_n) \quad \text{..........(2.22)}$$

$$E(X_1) = \sum_{x1} \sum_{x2} \ldots \sum_{xn} x_1.\, f(x_1,x_2,\ldots,x_n) \quad \text{.........................(2.23)}$$

$$E(X_i) = \sum_{x1} \sum_{x2} \ldots \sum_{xn} x_i.\, f(x_1,x_2,\ldots,x_n) \text{ เมื่อ } i = 1,\ldots, n \quad \text{......(2.24)}$$

2.10.4.2 กรณีที่ $X_1, X_2, \ldots\ldots, X_n$ เป็นตัวแปรสุ่มต่อเนื่อง

$$E(Z) = \int_{x1} \int_{x2} \ldots \int_{xn} u(x_1,x_2,\ldots,x_n)\, f(x_1,x_2,\ldots,x_n)\, dx_n\, dx_{n-1} \ldots dx_1 \quad \text{...(2.25)}$$

$$E(X_1) = \int_{x1} \int_{x2} \ldots \int_{xn} x_1\, f(x_1,x_2,\ldots,x_n)\, dx_n\, dx_{n-1} \ldots dx_1 \quad \text{...............(2.26)}$$

$$E(X_i) = \int_{x1} \int_{x2} \ldots \int_{xn} x_i\, f(x_1,x_2,\ldots,x_n)\, dx_n\, dx_{n-1} \ldots dx_1$$

เมื่อ $i = 1,\ldots, n$ ....(2.27)

ตัวอย่างที่ 2.18 X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่องมีการแจกแจงความน่าจะเป็นร่วมกันดังตาราง

	f(x,y)	
y \ x	0	1
-1	0.3	-
0	-	0.2
1	0.4	0.1

จงหา E(XY)

$$E(XY) = \sum_y \sum_x x.y.f(x,y)$$

x	0	0	0	1	1	1	$\sum$
y	-1	0	1	-1	0	1	
f(x,y)	0.3	-	0.4	-	0.2	0.1	1.00
xy.f(x,y)	0	0	0	0	0	0.1	0.1

$$\therefore E(XY) = 0.1$$

ตัวอย่างที่ 2.19 X และ Y เป็นตัวแปรสุ่มต่อเนื่องมีฟังค์ชั่นการแจกแจงความน่าจะเป็นร่วมกันดังนี้

$$f(x,y) = \frac{x(1+3y^2)}{4} \quad \text{เมื่อ } 0 \le x < 2 \text{ และ } 0 \le y < 1$$

จงหา $E(\frac{Y}{X})$

$$E(\frac{Y}{X}) = \int_0^1 \int_0^2 \frac{y}{x} [\frac{x(1+3y^2)}{4}]\, dx\, dy$$

$$= \int_0^1 \int_0^2 \frac{1}{4} y (1+3y^2)\, dx\, dy$$

$$= \int_0^1 \frac{1}{4} y (1+3y^2)\, x\Big|_0^2\, dy$$

$$= \int_0^1 \frac{1}{2} y\,(1+3y^2)\ dy$$

$$= \frac{1}{2}\left(\frac{y^2}{2}+\frac{3y^4}{4}\right)\Big|_0^1$$

$$= \frac{1}{2}\left(\frac{1}{2}+\frac{3}{4}\right) = \frac{5}{8}$$

ตัวอย่างที่ 2.20 X, Y เป็นตัวแปรสุ่มต่อเนื่องมี

$f(x,y) = x + y$ เมื่อ $0 \le x < 1$ และ $0 \le y < 1$

จงพิสูจน์ว่า X และ Y เป็นอิสระต่อกัน

$$E(XY) = E(X)\,E(Y)$$

$$E(X) = \int_0^1 \int_0^1 x\, f(x,y)\, dx\, dy$$

$$= \int_0^1 \int_0^1 x\,(x+y)\, dx\, dy$$

$$= \int_0^1 \left(\frac{x^3}{3}+\frac{x^2}{2}y\right)\Big|_0^1\ dy$$

$$= \int_0^1 \left(\frac{1}{3}+\frac{1}{2}y\right)\ dy$$

$$= \left(\frac{1}{3}y+\frac{1}{2}\ \frac{y^2}{2}\right)_0^1$$

$$= \frac{1}{3}+\frac{1}{4} = \frac{7}{12}$$

$$E(Y) = \int_0^1 \int_0^1 y\, f(x,y)\, dx\, dy$$

$$= \int_0^1 \int_0^1 y\,(x+y)\, dx\, dy$$

$$= \int_0^1 \left(y\frac{x^2}{2}+y^2x\right)\Big|_0^1\ dy$$

$$= \int_0^1 \left(\frac{y}{2}+y^2\right)\ dy$$

$$= \left(\frac{y^2}{4}+\frac{y^3}{3}\right)_0^1$$

$$= \left(\frac{1}{4}+\frac{1}{3}\right) = \frac{7}{12}$$

$$E(XY) = \int_0^1 \int_0^1 xy\, f(x,y)\, dx\, dy$$

$$= \int_0^1 \int_0^1 xy\,(x+y)\, dx\, dy$$

$$= \int_0^1 \left(y\frac{x^3}{3}+y^2\frac{x^2}{2}\right)_0^1\ dy$$

$$= \int_0^1 \left(\frac{y}{3}+\frac{y^2}{2}\right)\ dy$$

$$= \left(\frac{y^2}{6}+\frac{y^3}{6}\right)_0^1$$

$$= \frac{1}{6}+\frac{1}{6} = \frac{1}{3}$$

$E(X\,Y) \neq E(X).\,E(Y)$

แสดงว่า X, Y ไม่เป็นอิสระต่อกัน

2.10.5 การหาความแปรปรวน (Variance) ของตัวแปรสุ่มและฟังค์ชั่นของตัวแปรสุ่ม

ความแปรปรวน (Variance) คือพารามิเตอร์ของตัวแปรสุ่ม (X) หรือของฟังค์ชั่นการแจกแจงความน่าจะเป็น (f(x)) ซึ่งแสดงรูปร่างของการแจกแจงความน่าจะเป็น สามารถคำนวณหาได้โดยใช้ Expected Value

ให้ $Var(X) = \sigma^2$ = ความแปรปรวนของ X

$Var(X) = \sigma^2 = E(X-\mu)^2$

$= \sum_{x} (x-\mu)^2 \,.\, f(x)$ .... Discrete

$= \int_{x} (x-\mu)^2 \,.\, f(x)\, dx$ .... Continuous......(2.28)

= Moment ที่ 2 รอบ μ

σ คือ ส่วนเบี่ยงเบนมาตราฐาน (Standard Deviation) ของ X

ถ้า Z = u(X) = ฟังค์ชั่นใด ๆ ของ X

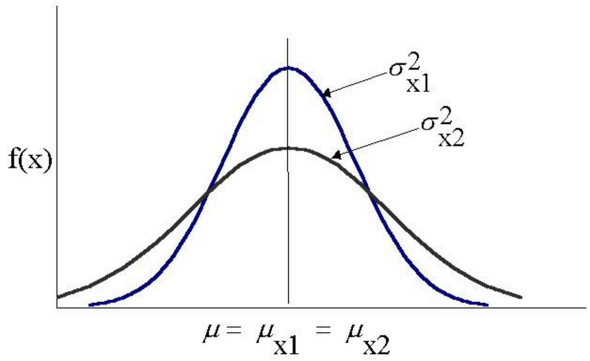
$Var(Z) = Var(u(X)) = \sigma^2_{u(x)}$

$= E(u(X)-\mu_{U(X)})^2$

สมมติ X_1 มีการแจกแจงตามฟังค์ชั่น $f(x_1)$ มีค่าเฉลี่ย μ_{x1} และความแปรปรวน σ_{x1}^2

X_2 มีการแจกแจงตามฟังค์ชั่น $f(x_2)$ มีค่าเฉลี่ย μ_{x2} และความแปรปรวน σ_{x2}^2

ถ้า $\mu_{x1} = \mu_{x2}$ และ $\sigma_{x1}^2 > \sigma_{x2}^2$ ความแปรปรวนจะมีผลต่อการแจกแจงดังรูปที่ 2.4



รูปที่ 2.4 ผลของค่าความแปรปรวนต่อรูปร่างของการแจกแจงความน่าจะเป็น

ตัวอย่างที่ 2.21 ให้ตัวแปรสุ่ม X เป็นจำนวนของผลิตภัณฑ์ชำรุดที่ตรวจสอบพบจากการสุ่มตัวอย่างผลิตภัณฑ์ 3 ชิ้น จากสายการผลิต และมีการแจกแจงความน่าจะเป็น ดังนี้

x	0	1	2	3
f(x)	0.51	0.38	0.10	0.01

จงคำนวณหาค่าความแปรปรวนของตัวแปรสุ่ม X

X	0	1	2	3	$\sum$
f(x)	0.51	0.38	0.10	0.01	1.00
x.f(x)	0	0.38	0.2	0.03	0.61 = μ_x
x^2.f(x)	0	0.38	0.4	0.09	0.87 = $E(X^2)$

$$\sigma^2 = E(X^2) - \mu_x^2$$
$$= 0.87 - (0.61)^2$$
$$= 0.4979$$

ตัวอย่างที่ 2.22 ให้ฟังค์ชั่นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง X คือ

$$f(x) = 2x^2 + \frac{1}{3} \qquad \text{เมื่อ } 0 \le x < 1$$

จงหาค่าเฉลี่ยและค่าความแปรปรวนของ X

$$\mu = E(X) = \int_0^1 x\left(2x^2 + \frac{1}{3}\right) dx$$

$$= \left(\frac{x^4}{2}+\frac{x^2}{6}\right)\Big|_0^1$$

$$= \frac{2}{3}$$

$$E(X^2) = \int_0^1 x^2(2x^2+\frac{1}{3}).\,Dx$$

$$= \frac{23}{45}$$

$$\sigma^2 = E(X^2)-\mu^2$$

$$= \frac{23}{45}-\left(\frac{2}{3}\right)^2$$

$$= \frac{1}{15}$$

ตัวอย่างที่ 2.23 ให้ $u(X) = 2X+3$ เมื่อ X เป็นตัวแปรสุ่ม ซึ่งมีการแจกแจงความน่าจะเป็นดังนี้

x	0	1	2	3
f(x)	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

จงหาความแปรปรวนของ u(X)

X	0	1	2	3	$\sum$
f(x)	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	1.00
x.f(x)	0	$\frac{1}{8}$	1	$\frac{3}{8}$	$1.5 = \mu_x$
x^2.f(x)	0	$\frac{1}{8}$	2	$\frac{9}{8}$	$\frac{26}{8} = E(X^2)$
U(x)=2x+3	3	5	7	9	
u(x).f(x)	$\frac{3}{4}$	$\frac{5}{8}$	$\frac{7}{2}$	$\frac{9}{8}$	$6 = \mu_{u(x)}$
$(u(x)-\mu_{u(x)})^2$	9	1	1	9	
$(u(x)-\mu_{u(x)})^2 f(x)$	$\frac{9}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{9}{8}$	$4 = \sigma^2_{u(x)}$

หรือ $\quad Var(u(X)) = E[(2X+3)-6]^2$

$$= E(2X-3)^2$$

$$= E(4X^2-12X+9)$$

$= \quad 4E(X^2)-12E(X)+9$

$= \quad 4 \times \frac{26}{8} -12 \times 1.5 + 9$

$= \quad 13 - 18 + 9 \;=\; 4$

ตัวอย่างที่ 2.24 ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นดังนี้

$f(x) = \frac{x^2}{3}$ เมื่อ $-1 \le x < 2$

จงหาค่าความแปรปรวนของ $u(x) = 4x + 3$

$E(4X+3) = \int_{-1}^{2} (4x+3)\left(\frac{x^2}{3}\right) dx$

$= \frac{1}{3}\int_{-1}^{2} (4x^3+3x^2)\, dx$

$= 8$

$\sigma^2_{(4x+3)} = E\,[(4X+3)- \mu_{(4x+3)}]^2$

$= E\,[(4X+3-8]^2$

$= \int_{-1}^{2} (4x-5)^2 . \frac{x^2}{3} .\, dx$

$= \frac{1}{3}\int_{-1}^{2} (16x^4-40x^3+25x^2).\, dx$

$= 10\frac{1}{5}$

2.10.6 กฎเกี่ยวกับ Variance

(1) $Z = aX+b$

$\sigma^2_z = \sigma^2_{ax+b} = a^2\,\sigma^2$ .... (2.29)

(2) $Z = aX+bY$ เมื่อ X, Y คือตัวแปรสุ่ม 2 ตัว

$\sigma^2_z = a^2\sigma^2_x + b^2\sigma^2_y+2ab\sigma_{xy}$ .... (2.30)

เมื่อ σ_{xy} = Covariance ของ X และ Y

(3) $Z = a_1 X_1+a_2 X_2 +..........+a_nX_n$ เมื่อ $X_1, X_2,....X_n$ คือตัวแปรสุ่ม

$\sigma^2_z = a_1^{\,2}\,\sigma^2_{x1} +a^2_{\,2}\sigma^2_{x2} + ...+ a_n^{\,2}\,\sigma^2_{xn}$........... (2.31)

ถ้า $X_1\ X_2$,.........., X_n independent

Covariance จะมีค่าเป็น 0

ตัวอย่างที่ 2.25 ให้ X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันมี $\sigma^2_X = 5, \sigma^2_Y = 3$

จงหาค่า σ^2_Z เมื่อ $Z = X + 4Y - 3$

$$Z = X + 4Y - 3$$
$$\sigma^2_Z = \sigma^2_X + 16\sigma^2_Y + 0$$
$$= 5 + 48$$
$$= 53$$

ตัวอย่างที่ 2.26 ให้ X และ Y เป็นตัวแปรสุ่มที่ไม่เป็นอิสระต่อกันมี $\sigma^2_X = 5, \sigma^2_Y = 3$ และ $\sigma_{XY} = 1$ จงหา σ^2_Z เมื่อ $Z = 2X - 3Y + 5$

$$Z = 2X - 3Y + 5$$
$$\sigma^2_Z = 4\sigma^2_X + 9\sigma^2_Y - (2x2x3)\,\sigma_{XY}$$
$$= 20 + 27 - 12$$
$$= 35$$

2.10.7 ความแปรปรวนร่วมกันของตัวแปรสุ่ม (Covariance)

ให้ X และ Y เป็นตัวแปรสุ่มมีฟังก์ชั่นความน่าจะเป็นร่วมกันเท่ากับ f(x,y) และมีค่าเฉลี่ยเท่ากับ μ_X และ μ_Y ตามลำดับ

ให้ Cov (X,Y) = σ_{XY} = ความแปรปรวนร่วมกันของตัวแปรสุ่ม X และ Y

$$\text{Cov}(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \sum_y \sum_x (x-\mu_X)(y-\mu_Y)\, f(x,y) \quad \text{……........Discrete}$$
$$= \int_y \int_x (x-\mu_X)(y-\mu_Y)\, f(x,y)\, dx\, dy \quad \text{....Continuous...(2.32)}$$

หรือ $E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X.\mu_Y$ ....………...(2.33)

ถ้า X, Y = independent

$$E(XY) = E(X)\,E(Y) = \mu_X.\,\mu_Y$$
$$\sigma_{XY} = 0$$

ตัวอย่างที่ 2.27 ถ้าจำนวนลูกบอลในกล่องมีจำนวน 8 ลูก เป็นสีดำ 3 ลูก สีแดง 2 ลูก และสีเขียว 3 ลูกถ้าหยิบลูกบอลอย่างสุ่ม 2 ลูก ให้ X, Y เป็นตัวแปรสุ่มซึ่งแสดงถึงจำนวนลูกบอลสีดำและสีแดงที่หยิบขึ้นมา ซึ่งมีความน่าจะเป็นร่วมกันดังนี้

f(x,y)					h(y)
		X			
		0	1	2	
Y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	-	$\frac{3}{7}$
	2	$\frac{1}{28}$	-	-	$\frac{1}{28}$
g(x)		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

จงหาค่าความแปรปรวนร่วมกันของตัวแปรสุ่ม X, Y

X	0	0	0	1	1	1	2	2	2	$\sum$
Y	0	1	2	0	1	2	0	1	2	
f(x,y)	$\frac{3}{28}$	$\frac{3}{14}$	$\frac{1}{28}$	$\frac{9}{28}$	$\frac{3}{14}$	-	$\frac{3}{28}$	-	-	1.00
xy.f(x,y)	0	0	0	0	$\frac{3}{14}$	-	-	-	-	$\frac{3}{14}$=E(XY)

x	0	1	2	$\sum$
g(x)	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1.00
x.g(x)	0	$\frac{15}{28}$	$\frac{6}{28}$	$\frac{3}{4}$ = E(X)

y	0	1	2	$\sum$
h(y)	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$	1.00
y.h(y)	0	$\frac{3}{7}$	$\frac{2}{28}$	$\frac{1}{2}$ = E(Y)

$$\text{Cov (XY)} = \sigma_{XY} = E(XY)-E(X)E(Y)$$

$$= \frac{3}{14}-\frac{3}{4}\text{x}\frac{1}{2} = -\frac{9}{56}$$

ตัวอย่างที่ 2.28 ให้ตัวแปรสุ่ม X, Y มีฟังก์ชั่นการแจกแจงความน่าจะเป็นร่วมกันคือ

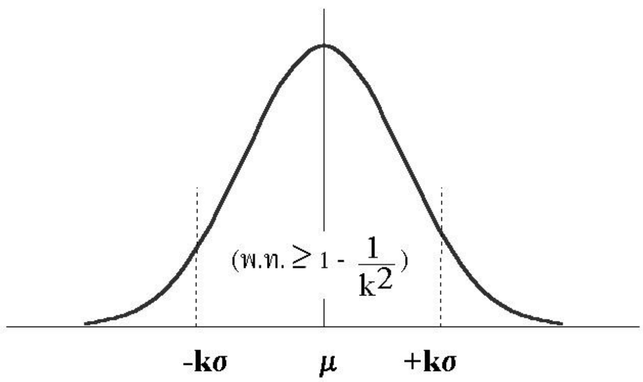
$f(x,y) = 8xy$ เมื่อ $0 \le x < 1$

จงหาค่าความแปรปรวนร่วมกันของ X และ Y

$$g(x) = \int_0^x 8xy\,.\,dy$$
$$= 4xy^2 \Big|_0^x$$
$$= 4xy^3 \quad \text{เมื่อ} \quad 0 \le x < 1$$
$$h(y) = \int_y^1 8xy\,.\,dx$$
$$= 4x^2y \Big|_y^1$$
$$= 4y(1-y^2) \text{เมื่อ} \quad 0 \le y < 1$$
$$\mu_X = E(X) = \int_0^1 x.\,4x^3\,.\,dx = \frac{4}{5}$$
$$\mu_Y = E(Y) = \int_0^1 y.\,4y(1-y^2).\,Dy = \frac{8}{15}$$
$$E(XY) = \int_0^1 \int_y^1 xy.\,8xy.\,dx.\,Dy = \frac{4}{9}$$
$$\sigma_{xy} = E(XY) - E(X).\,E(Y)$$
$$= \frac{4}{9} - \left(\frac{4}{5} \times \frac{8}{15}\right) = \frac{4}{225}$$

2.10.8 ทฤษฎี Chebyshev

นักคณิตศาสตร์ชาวรัสเซียชื่อ P.L. Chebyshev พบว่า



$$P(\mu - k\sigma \le X < \mu + k\sigma) \ge 1 - \frac{1}{k^2}$$

พิสูจน์

$$\sigma^2 = E\,(X-\mu)^2$$
$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\,dx$$

$$+ \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx$$

ถ้า $|x-\mu| = k\sigma$

$(x-\mu)^2 = k^2\sigma^2$

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx$$

$$\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \leq \frac{1}{k^2}$$

เนื่องจาก $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\therefore \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)\,dx = 1 - \int_{-\infty}^{\mu-k\sigma} f(x)\,dx - \int_{\mu+k\sigma}^{\infty} f(x)\,dx$$

$$\geq 1-\frac{1}{k^2}$$

$$\therefore P(\mu-k\sigma<X<\mu+k\sigma) \geq 1-\frac{1}{k^2}$$

ถ้า k = 2

$$P(X-2\sigma<X<\mu+2\sigma) \geq 1-\frac{1}{2^2}$$

$$\geq \frac{3}{4}$$

ตัวอย่างที่ 2.29 ให้ X เป็นตัวแปรสุ่มมี $\mu = 8$ และ $\sigma^2 = 9$ ไม่ทราบการแจกแจง

จงหา

(ก) $P(-4 < X < 20)$

(ข) $P(|X-8| \geq 6)$

(ก) P(-4 < X < 20)

$= P[(8-4x3) < X < (8+4x3)]$

$k = 4$

$$P(-4 < X < 20) \geq 1-\frac{1}{k^2}$$

$$\geq 1-\frac{1}{4^2}$$

$$\geq 1-\frac{1}{16} \quad (=\frac{15}{16})$$

(ข) $P(|X-8| \geq 6)$

$$= P(|X-8| \geq 2x3)$$

$$= 1 - P[(8-2x3) < X < (8+2x3)]$$

$$P(|X-8| \geq 6) \leq 1-(1-\frac{1}{k^2})$$

$$\leq \frac{1}{k^2}$$

$$\leq \frac{1}{2^2}$$

$$\leq \frac{1}{4}$$

2.11 แบบฝึกหัด

1. ในการทดลองโยนลูกเต๋า 2 ลูก ให้ X คือตัวแปรสุ่มแสดงค่า Absolute Value ของผลต่างของแต้มลูกเต๋า 2 ลูก จงหาตารางการแจกแจงความน่าจะเป็นของ X พร้อมเขียนกราฟ และหาสมการ Probability Mass Function, f(x)

2. ในการทดลองโยนลูกบอล 5 ลูก ลงกล่อง 4 กล่อง กำหนดว่าความน่าจะเป็นที่ลูกบอลจะลงในกล่องใดกล่องหนึ่งมีค่าเท่ากัน ให้ X คือจำนวนลูกบอลที่ตกลงในกล่อง 3 จงหา

    (ก) X คือตัวแปรสุ่มแบบไหน

    (ข) การแจกแจงความน่าจะเป็นของ X (เขียน Tree Diagram ประกอบ)

    (ค) ความน่าจะเป็นที่บอลจะตกลงในกล่อง 3 ทั้ง 5 ลูก

3. กล่องใบหนึ่งบรรจุลูกบอล 10 ลูก เป็นสีดำ 3 ลูก สีขาว 2 ลูก และสีแดง 5 ลูก สุ่มหยิบลูกบอลทีละลูก 2 ครั้ง

    (ก) ให้ X คือจำนวนลูกบอลสีขาวที่หยิบได้

    จงหาการแจกแจงความน่าจะเป็นของ X

    (ข) ให้ X คือจำนวนลูกบอลสีขาวที่หยิบได้ Y คือจำนวนลูกบอลสีดำที่หยิบได้

    จงหาการแจกแจงความน่าจะเป็นร่วมของ X และ Y

    (ค) จากข้อ (ข) ถ้าให้ Z คือจำนวนลูกบอลสีแดงที่หยิบได้

    จงหาความน่าจะเป็นที่จะหยิบได้ลูกบอลสีแดงทั้ง 2 ลูก

    เขียน Tree Diagram เปรียบเทียบเหตุการณ์ในข้อ (ก), (ข) และ (ค)

4. ในการทดลองโยนเหรียญอันหนึ่ง 3 ครั้ง ถ้ากำหนดให้

    X = จำนวนครั้งของการเกิดหัวในการโยนเหรียญครั้งแรก

    Y = จำนวนครั้งของการเกิดหัวในการโยนเหรียญ 3 ครั้ง

    จงหา

    (ก) Marginal Distribution ของ X และฟังค์ชั่นของ g(x)

    (ข) Marginal Distribution ของ Y และฟังค์ชั่นของ h(y)

    (ค) การแจกแจงความน่าจะเป็นร่วมของ X และ Y และฟังค์ชั่นของ f(x,y)

    (ง) Conditional Distribution ของ Y|X และฟังค์ชั่นของ f(y|x)

    (จ) Conditional Distribution ของ X|Y และฟังค์ชั่นของ f(x|y)

5. กำหนดให้ตัวแปรสุ่ม X ซึ่งมีฟังก์ชั่นการแจกแจงความน่าจะเป็นดังนี้

$f(x) = kx^2$ เมื่อ $|x| \leq k$

จงหา

(ก) ค่าของ k

(ข) $P(0 \leq X < 0.5)$

6. กำหนดให้ฟังก์ชั่นการแจกแจงความน่าจะเป็นร่วมของ X และ Y คือ

$f(x,y) = 8xy$ เมื่อ $0 \leq x < y < 2$

จงหา

(ก) Marginal Distribution ของ X และ Y

(ข) ตรวจสอบว่า X และ Y เป็นอิสระต่อกัน

(ค) $P(\frac{1}{4} \leq X < 1 \mid Y = \frac{3}{2})$

7. เครื่องบินทิ้งระเบิดลำหนึ่งมีภาระกิจเพื่อไปทิ้งระเบิดทำลายรางรถไฟแห่งหนึ่ง ถ้าเครื่องบินลำนี้ทิ้งระเบิดลูกใหญ่ขณะที่บินสูง 40 เมตร เหนือรางรถไฟ หรือทิ้งระเบิดลูกเล็ก ขณะที่บินสูง 15 เมตร เหนือรางรถไฟ จะสามารถทำสายรางรถไฟได้

ถ้าให้ X = ระยะความสูงของเครื่องบินเหนือรางรถไฟขณะทิ้งระเบิด

$f(x) = \frac{100 - x}{5,000}$ เมื่อ $0 \leq x < 100$

จงหา

(ก) ความน่าจะเป็นที่จะสามารถทำลายรางรถไฟได้ถ้าทิ้งระเบิดลูกใหญ่ 1 ลูก

(ข) ความน่าจะเป็นที่จะสามารถทำลายรางรถไฟได้ถ้าทิ้งระเบิดลูกเล็ก 1 ลูก

(ค) ความน่าจะเป็นที่จะสามารถทำลายรางรถไฟได้ถ้าทิ้งระเบิดลูกใหญ่ 5 ลูก และลูกเล็ก 5 ลูก

8. ในการทดลองสุ่มหยิบผลไม้ 400 ผล จากถุงซึ่งบรรจุผลไม้ 3 ชนิด คือ ส้ม 4 ผล มะม่วง 2 ผล และมังคุด 3 ผล ถ้าให้

X = จำนวนส้มที่หยิบได้

Y = จำนวนมะม่วงที่หยิบได้

จงหา (ก) การแจกแจงความน่าจะเป็นร่วมของ X และ Y

(ข) $P[(X, Y) \varepsilon A]$ เมื่อ $A = \{ (x, y) \mid |x-y| \leq 1 \}$

9. คะแนนสอบไล่ของวิชาความน่าจะเป็นและสถิติประยุกต์สำหรับวิศวกร ซึ่งมีนิสิต 60 คน มีค่าดังตาราง

23	60	79	32	57	74	52	70	82	36
80	77	81	95	41	65	92	85	55	76
52	10	64	75	78	25	80	98	81	67
41	71	83	54	64	72	88	62	74	43
60	78	89	76	84	48	84	90	15	79
34	67	17	82	69	74	63	80	85	61

(ก) จงสร้าง Stem and Leaf ของคะแนนสอบไล่ดังกล่าว

(ข) จงหาการแจกแจงความถี่สัมพัทธ์

(ค) สร้างกราฟแจกแจงความถี่สัมพัทธ์ (Relative Frequency Histogram) พร้อมลากเส้นประมาณของ f(x) และบอกว่ากราฟบิดเบี้ยวอย่างไร

(ง) สร้างกราฟการแจกแจงความถี่สัมพัทธ์สะสม พร้อมลากเส้นประมาณของ F(x)

(จ) จงหา First Quartile และ Seventh Decile

10. กำหนดฟังค์ชั่นการแจกแจงความน่าจะเป็นร่วมกันของ X และ Y คือ

$$f(x, y) = \frac{x+y}{30} \quad \text{เมื่อ} \quad x = 0, 1, 2, 3 \quad \text{และ} \quad y = 0, 1, 2$$

จงหา

(ก) $P(X \leq 2, Y = 1)$

(ข) $P(X > 2, Y \leq 1)$

(ค) $P(X > Y)$

(ง) $P(X+Y = 4)$

11. ในการทดลองทอดลูกเต๋า 1 ลูก ถ้ากำหนดให้ X = แต้มลูกเต๋าที่ทอดได้

จงหา

(ก) (X) และ VAR (X)

(ข) E (Y) เมื่อ $Y = 3X^2 + 6$

(ค) VAR (Z) เมื่อ $Z = 2X - 5$

12. ให้ X และ Y เป็นตัวแปรสุ่มซึ่งมีฟังก์ชั่นความน่าจะเป็นคือ

$f(x) = \frac{x}{3}$ เมื่อ $x = 1, 2$

และ $f(y|x) = \binom{x}{y}0.5^{x}$ เมื่อ $y = 0, 1, \ldots\ldots, x$

$x = 1, 2$

จงหา

(ก) E (X) และ VAR (X)

(ข) E (Y) และ E (XY)

13. ตัวแปรสุ่ม X และ Y เป็นอิสระต่อกัน

ถ้า $E(X) = E(Y) = 10$ แต่ $\sigma_x \neq \sigma_y$

กำหนดให้ $Z = aX + (1-a)Y$ เมื่อ $0 \leq a < 1$

จงหา

(ก) E (Z)

(ข) ค่า a ในเทอมของ σ_x^2 และ σ_y^2 ซึ่งทำให้ VAR (Z) มีค่าต่ำที่สุด

14. ในการทำฟาร์มไก่ สมมติให้ไก่แต่ละตัวที่เกิดมาสามารถทำกำไรให้เจ้าของฟาร์มได้ตัวละ X บาท ถ้า X คือตัวแปรสุ่มต่อเนื่องซึ่งมีฟังก์ชั่นความน่าจะเป็นคือ

$f(x) = \frac{2}{9}(x+1)$ เมื่อ $-1 \leq x < 2$

จงหา

(ก) กำไรเฉลี่ยต่อลูกไก่ 1 ตัว

(ข) ความแปรปรวนของผลกำไรต่อลูกไก่ 1 ตัว

15. โรงงานทำธงชาติแห่งหนึ่ง ทำการตัดผ้าสีแดง สีขาว และสีน้ำเงินเพื่อเย็บธงชาติ

กำหนดให้ X = ความกว้างของผ้าสีแดงที่ตัดได้มี $\mu_x = 10$ ซม. และ $\sigma_x = 1.25$ ซม.

Y = ความกว้างของผ้าสีขาวที่ตัดได้มี $\mu_y = 10$ ซม. และ $\sigma_y = 0.75$ ซม.

Z = ความกว้างของผ้าสีน้ำเงินที่ตัดได้มี $\mu_z = 15$ ซม. และ $\sigma_z = 1$ ซม.

ถ้ากำหนดให้ W คือความกว้างของธงชาติซึ่งเท่ากับ 2X+2Y-Z

จงหา E(W) และ Var (W)

16. เกมดึงไพ่ออกจากสำรับ 1 ใบ แบบใส่คืนมีกติกาดังนี้

ได้ไพ่	ได้รับเงิน (บาท)
J	10
Q	20
K หรือ A	50
2-9	-

ถ้าในการเล่นแต่ละครั้งผู้เล่นต้องจ่ายเงินครั้งละ 10 บาท ถามว่าเกมนี้ยุติธรรมหรือไม่

17. ให้ COV (X, Y) คือ Covariance ของ X และ Y

จงพิสูจน์ว่า Cov (aX, bY) = ab Cov (X, Y)

18. ให้ X และ Y คือตัวแปรสุ่ม 2 ตัว และ

$$Z = aX + bY$$

จงพิสูจน์ว่า $\sigma_Z^2 = a^2 \sigma_X{}^2 + b^2 \sigma_y{}^2 + 2ab\sigma_{xy}$

และถ้า X และ Y เป็นอิสระต่อกัน จงพิสูจน์ว่า $\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2$

19. กำหนดว่าตัวแปรสุ่ม X คือ อายุการใช้งาน (ชม.) ของ Compressor ซึ่งมี Density Function คือ

$$f(x) = \frac{1}{900} e^{-x/900} \quad \text{เมื่อ } x > 0$$

จงหา

(ก) อายุการใช้งานเฉลี่ยของ Compressor

(ข) ค่าความแปรปรวนของ X

(ค) $E(X+5)^2$

20. ให้ X คือค่าที่ได้จากการทอดลูกเต๋าสีเขียว

Y คือค่าที่ได้จากการทอดลูกเต๋าสีแดง

จงหาค่าความแปรปรวนของ

(ก) 2X-Y

(ข) X+3Y-5

21. บริษัทผลิตสายโทรศัพท์แห่งหนึ่งกำหนดว่าค่าเฉลี่ยของสายโทรศัพท์ยาว 130 ซม. และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 15 ซม. จงคำนวณหาว่ามีสายโทรศัพท์กี่เปอร์เซ็นต์ที่ยาวกว่า 175 ซม.

บทที่ 3
การแจกแจงความน่าจะเป็น
(Probability Distributions)

3.1 คำนำ

ตัวแปรสุ่มคือ ตัวแปรที่แสดงผลลัพธ์ของการทดลองที่เราต้องการทราบ มีค่าเป็นตัวเลข ความรู้เกี่ยวกับการแจกแจงความน่าจะเป็นของตัวแปรสุ่มจะบอกให้เราทราบว่า

- ตัวแปรสุ่ม X มีค่าเป็นอะไรได้บ้าง
- f(x) และ F(x) มีลักษณะอย่างไร มีพารามิเตอร์อะไรบ้าง และหาได้อย่างไร
- พารามิเตอร์ของ X เช่น μ และ σ^2 มีค่าเท่าไร
- $P(X \le x)$ มีค่าเท่าไร ถ้าทราบค่า x
- x มีค่าเท่าไร ถ้าทราบค่า $P(X \le x)$

ตัวแปรสุ่มแบ่งออกได้เป็น 2 ประเภทคือ ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable) และตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable) ในบทนี้จะกล่าวถึงการแจกแจงความน่าจะเป็นของทั้งตัวแปรสุ่มไม่ต่อเนื่อง และตัวแปรสุ่มต่อเนื่อง

3.2 ตัวแปรสุ่มไม่ต่อเนื่อง

ตัวแปรสุ่มไม่ต่อเนื่องได้แก่ตัวแปรสุ่มที่แสดงค่าเป็นเลขจำนวนเต็ม เช่น จำนวนคน จำนวนเครื่องคอมพิวเตอร์ จำนวนรถยนต์ จำนวนชิ้นส่วน เครื่องจักร เครื่องมือต่าง ๆ เป็นต้น ในบทนี้จะกล่าวถึงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง 6 ชนิดคือ

(1) การแจกแจงแบบเอกรูป หรือแบบสม่ำเสมอ (Uniform Distribution)
(2) การแจกแจงแบบทวินาม (Binomial Distribution)
(3) การแจกแจงแบบพหุนาม (Multinomial Distribution)
(4) การแจกแจงแบบไฮเปอร์จีออเมตริก (Hypergeometric Distribution)
(5) การแจกแจงแบบปัวซอง (Poisson Distribution)
(6) การแจกแจงแบบทวินามลบและการแจกแจงแบบเรขาคณิต (Negative Binomial and Geometric Distribution)

3.2.1 การแจกแจงแบบเอกรูปหรือแบบสม่ำเสมอ (Uniform Distribution)

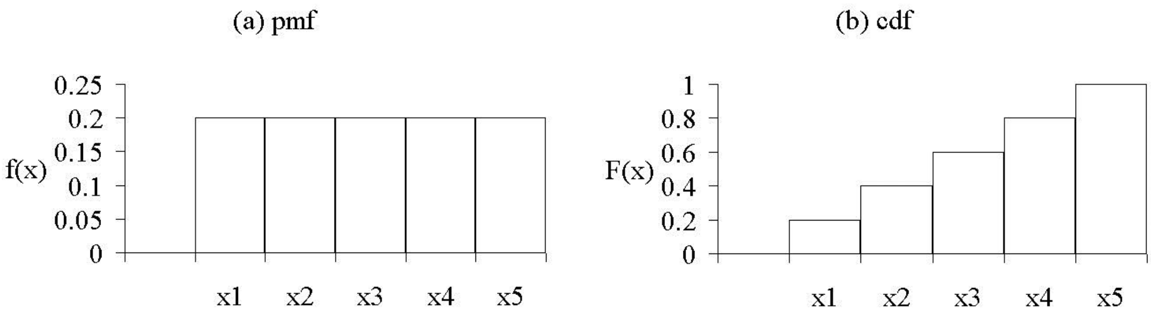
การแจกแจงแบบเอกรูปหรือสม่ำเสมอ เกี่ยวข้องกับการทดลองซึ่งโอกาสการเกิดผลลัพธ์ใด ๆ มีค่าเท่ากัน เช่น การทอดลูกเต๋า การหยิบไพ่ การจับสลากล๊อตเตอรี่ เป็นต้น

ให้ X = ตัวแปรสุ่มแบบเอกรูป

$$f(x) = u(x;k) = \frac{1}{k} \quad \text{เมื่อ } x = x_1, x_2, \ldots, x_i, \ldots, x_k \qquad \ldots\ldots(3.1)$$

$$F(x_i) = \sum_{j=1}^{i} f(x_j) = \sum_{j=1}^{i} \frac{1}{k} = \frac{i}{k} \qquad \ldots\ldots(3.2)$$

เมื่อ k คือ พารามิเตอร์ของการแจกแจงแบบเอกรูป



รูปที่ 3.1 กราฟการแจกแจงความน่าจะเป็นแบบเอกรูป

$$\mu = \sum_{i=1}^{k} x_i f(x_i)$$

$$= \frac{1}{k}\sum_{i=1}^{k} x_i \qquad \ldots\ldots(3.3)$$

$$\sigma^2 = \sum_{i=1}^{k} (x_i-\mu)^2 f(x_i)$$

$$= \frac{1}{k}\sum_{i=1}^{k} (x_i-\mu)^2 \qquad \ldots\ldots(3.4)$$

หรือ $$\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2-\mu^2 \qquad \ldots\ldots(3.5)$$

ตัวอย่างที่ 3.1 ให้ X เป็นตัวแปรสุ่มแบบเอกรูปแสดงผลของการทอดลูกเต๋า จงหา f(x), F(x), μ และ σ^2

$$f(x) = \frac{1}{6} \quad \text{เมื่อ } x = 1, 2, \ldots, 6$$

$$F(x) = \sum_{t=1}^{x} f(t) = \frac{x}{6} \quad \text{เมื่อ } x = 1, 2, \ldots\ldots\ldots, 6$$

$$\mu = \frac{1}{6}\sum_{i=1}^{6} x_i$$

$$= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

$$\sigma^2 = \frac{1}{6}\sum_{i=1}^{6} x_i^2 - \left(\frac{21}{6}\right)^2$$

$$= \frac{1}{6}(1+4+9+16+25+36) - \left(\frac{21}{6}\right)^2$$

$$= \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{105}{36} = 2.92$$

3.2.2 การแจกแจงแบบทวินาม (Binomial Distribution)

การทดลองแบบทวินาม (Binomial Process) มีคุณสมบัติที่สำคัญคือ

- เกี่ยวข้องกับการทดลอง n ครั้ง
- ผลลัพธ์การทดลองแต่ละครั้งมี 2 แบบ คือ
  - ความสำเร็จ (Success) หรือความล้มเหลว (Failure)
  - ถูกหรือผิด
  - ดีหรือเสีย
  - สีแดงหรือไม่ใช่สีแดง เป็นต้น
- การทดลองแต่ละครั้งเป็นอิสระต่อกัน
- ความน่าจะเป็นที่ผลการทดลองแต่ละครั้งเป็น "ความสำเร็จ" เท่ากับ p และ "ความล้มเหลว" เท่ากับ q (=1-p) และมีค่าคงที่

ให้ X = ตัวแปรสุ่มแบบทวินามแสดงจำนวนครั้งของความสำเร็จในการทดลอง n ครั้ง

$$f(x) = b(x; n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

เมื่อ x = 0, 1, 2, ..., n ……......…..(3.6)

$$F(r) = P(X \le r) = B(r; n,p)$$

$$= \sum_{x=0}^{r} b(x; n,p) = \sum_{x=0}^{r} \binom{n}{x} p^x q^{n-x}$$

$$= \binom{n}{0} q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2 q^{n-2} + \ldots\ldots\ldots + \binom{n}{r} p^r q^{n-r} \quad \ldots\ldots\ldots(3.7)$$

ตาราง Binomial Sum B(r; n,p) สำหรับค่า r, n, p ต่าง ๆ แสดงอยู่ในตารางที่ A1 ท้ายเล่ม

ถ้า x = n

$$F(n) = \sum_{x=0}^{n} b(x;\, n, p)$$

$$= (p+q)^n = 1.00$$

$$\mu = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!\,(x-1)!} p^{x-1} q^{n-x}$$

ให้ $y = x-1$

ถ้า $x = 1$; $y = 0$

$x = n$; $y = n-1$

$$\mu = np \sum_{y=0}^{n-1} \frac{(n-1)!}{(n-1-y)!\,y!} p^y q^{n-1-y}$$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y q^{n-1-y}$$

$$= np\,(p+q)^{n-1} = np \qquad \text{.....................................(3.8)}$$

$$E(X^2) = \sum_{x=0}^{n} x^2\, b(x;\, n,p)$$

$$= \sum_{x=0}^{n} \{x+x(x-1)\}\, b(x;\, n,p)$$

$$= \mu + \sum_{x=0}^{n} x(x-1) \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

$$= \mu + n(n-1)\,p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!\,(x-2)!} p^{x-2} q^{n-x}$$

ให้ $y = x-2$

ถ้า $x = 2$; $y = 0$

$x = n$; $y = n-2$

$$E(X^2) = \mu + n(n-1)\,p^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{(n-2-y)!\,y!} p^y q^{n-2-y}$$

$$= \mu + n(n-1)\,p^2\,(p+q)^{n-2}$$

$$= np + n(n-1)\,p^2$$

$$\sigma^2 = E(X^2) - \mu^2$$

$$= np + n(n-1)\,p^2 - (np)^2$$

$$= np + (np)^2 - np^2 - (np)^2$$

$$= np\,(1-p) = npq \qquad \text{.......................................(3.9)}$$

ตัวอย่างที่ 3.2 ในการทดลองสุ่มผลิตภัณฑ์ 3 ชิ้น จากขบวนการผลิต และความน่าจะเป็นที่จะสุ่มได้ผลิตภัณฑ์เสียเท่ากับ $\frac{1}{4}$ จงหา

(1) $P(X = 2)$ และ $P(X \leq 2)$

(2) μ, σ^2

(3) จงเขียน Tree Diagram แสดงผลลัพธ์ที่เป็นไปได้ของการทดลอง และความน่าจะเป็นของผลลัพธ์

วิธีทำ

n = 3

X = ตัวแปรสุ่มแบบทวินาม มีค่า 0, 1, 2, 3

(1) $P(X=2) = f(2)$

$$= \frac{3!}{(3-2)!\,2!}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$$

$$= 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{9}{64}$$

$P(X \leq 2) = F(2)$

$$= \sum_{x=0}^{2}\frac{3!}{(3-x)!\,x!}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{3-x}$$

$$= \left(\frac{3}{4}\right)^3 + 3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2 + 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$$

$$= \frac{27}{64}+\frac{27}{64}+\frac{9}{64} = \frac{63}{64}$$

หรือ $P(X \leq 2) = 1 - P(X = 3)$

$$= 1 - f(3)$$

$$= 1-\left(\frac{1}{4}\right)^3 = 1-\frac{1}{64} = \frac{63}{64}$$

(2) $\mu = np$

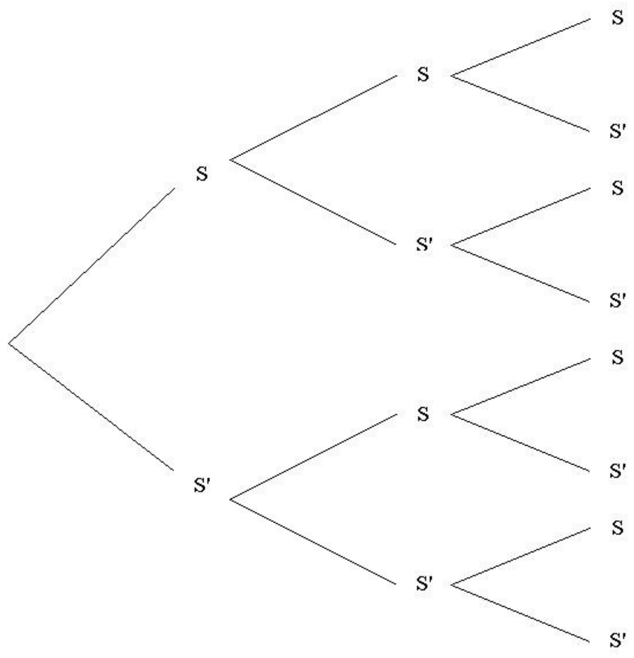
$$= 3x\frac{1}{4} = \frac{3}{4}$$

$\sigma^2 = npq$

$$= 3 \text{ x } \frac{1}{4} \text{ x } \frac{3}{4} = \frac{9}{16}$$

(3) Tree Diagram

ให้ S = ผลิตภัณฑ์เสียที่สุ่มได้ และ $P(S) = \frac{1}{4}$



$$\left(\frac{1}{4}\right)^3$$
$$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$$
$$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$$
$$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$$
$$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$$
$$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$$
$$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$$
$$\left(\frac{3}{4}\right)^3$$

x	f(x)
0	$\left(\frac{3}{4}\right)^3$
1	$3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$
2	$3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
3	$\left(\frac{1}{4}\right)^3$

ตัวอย่างที่ 3.3 จากตัวอย่างที่ 2.6 ถุงใบหนึ่งบรรจุผลไม้ 3 ชนิด คือ ส้ม 3 ผล มะม่วง 2 ผล และมังคุด 3 ผล ถ้าให้ X คือ จำนวนส้มที่หยิบได้จากการสุ่มเลือกผลไม้ 4 ครั้ง แบบใส่คืน จงหา f(x)

วิธีทำ

X = ตัวแปรสุ่มแบบทวินามแสดงจำนวนส้มที่หยิบได้

n = 4

x = 0, 1, 2, 3, 4

$p = \frac{3}{8}$

$$f(x) = \binom{4}{x}\left(\frac{3}{8}\right)^{x}\left(\frac{5}{8}\right)^{4-x}$$

x	0	1	2	3	4	$\sum$
f(x)	$\frac{625}{4096}$	$\frac{1500}{4096}$	$\frac{1350}{4096}$	$\frac{540}{4096}$	$\frac{81}{4096}$	1.0

3.2.3 การแจกแจงแบบพหุนาม (Multinomial Distribution)

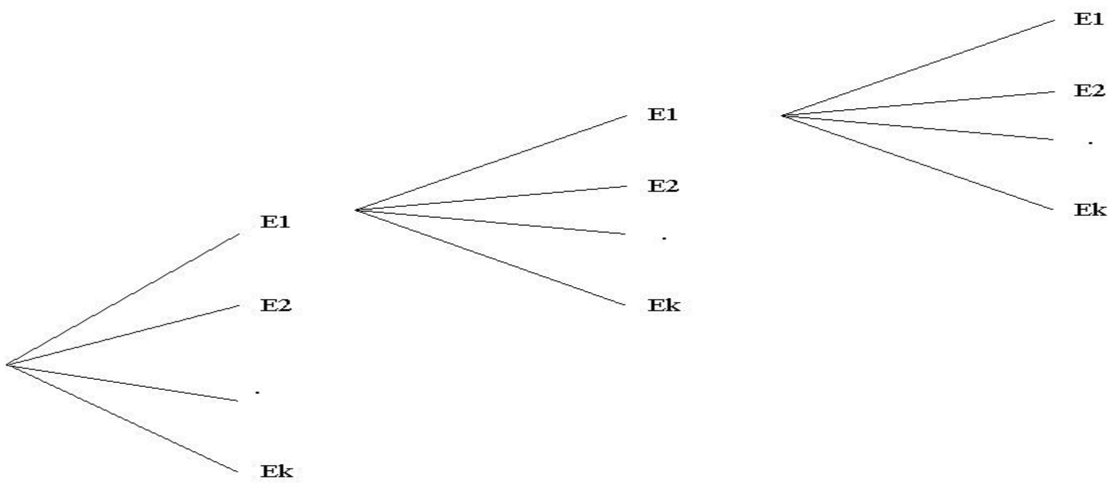
การทดลองแบบพหุนามคล้ายกับการทดลองแบบทวินาม คือ เกี่ยวข้องกับการทดลอง n ครั้ง การทดลองแต่ละครั้งเป็นอิสระต่อกันแต่ต่างกันตรงที่ผลลัพธ์ของการทดลองแต่ละครั้งมี k แบบ (ทวิ

นามมี 2 แบบ) ดังนั้นจึงอาจกล่าวได้ว่าการแจกแจงแบบทวินามเป็นกรณีพิเศษของการแจกแจงแบบพหุนามเมื่อ k = 2

ให้ $X_1, X_2, \ldots\ldots, X_k$ = คือเซทของตัวแปรสุ่มแบบพหุนามที่แสดงจำนวนครั้งของความสำเร็จในการทดลอง n ครั้ง

Outcomes	No.of Success	Prob. of Success
E_1	x_1	p_1
E_2	x_2	p_2
.	.	.
.	.	.
E_k	x_k	p_k
$\sum$	n	1.0

สามารถเขียน Tree Diagram แสดงผลลัพธ์ของการทดลองแบบพหุนามได้ดังนี้



$$f(x_1, x_2, \ldots, x_k ; p_1, p_2, \ldots, p_k, n) = \frac{n!}{x_1!\, x_2! \ldots x_k!} p_1^{x1} p_2^{x2} \ldots p_k^{xk}$$

เมื่อ $0 \leq x_i < n \quad (i = 1, \ldots, k)$

และ $\sum_{i=1}^{k} x_i = n$ ....(3.10)

ในทำนองเดียวกับกรณีการแจกแจงแบบทวินาม จะหา μ และ σ^2 ได้ดังนี้

ตัวแปรสุ่ม	μ	σ^2
X_1	np_1	$np_1(1-p_1)$
X_2	np_2	$np_2(1-p_2)$
.	.	.
.	.	.
X_k	np_k	$np_k(1-p_k)$

ตัวอย่างที่ 3.4 จากตัวอย่างที่ 3.3 ถ้าให้

X_1 = จำนวนส้มที่หยิบได้

X_2 = จำนวนมะม่วงที่หยิบได้

X_3 = จำนวนมังคุดที่หยิบได้

จากการสุ่มหยิบผลไม้ 4 ครั้ง แบบใส่คืน

จงหา

(1) p_1, p_2, p_3

(2) $P(X_1=1, X_2=1, X_3=2) = f(1, 1, 2)$

วิธีทำ

$n = 4, k = 3$

Outcomes	X	P(E.)
E_1 (ส้ม)	$0 \le x_1 < 4$	$p_1 = \frac{3}{8}$
E_2 (มะม่วง)	$0 \le x_2 < 4$	$p_2 = \frac{2}{8}$
E_3 (มังคุด)	$0 \le x_3 < 4$	$p_3 = \frac{3}{8}$
$\sum$	4	1.0

$P(X_1=1, X_2=1, X_3=2) = f(1, 1, 2)$

$$= f\left(1, 1, 2; \frac{3}{8}, \frac{2}{8}, \frac{3}{8}, 4\right)$$

$$= \frac{4\,!}{1\,!\,1\,!\,2\,!}\left(\frac{3}{8}\right)\left(\frac{2}{8}\right)\left(\frac{3}{8}\right)^2$$

$$= 12 \text{ x } \frac{3}{8} \text{ x } \frac{2}{8} \text{ x } \frac{9}{64} = \frac{81}{512}$$

3.2.4 การแจกแจงแบบไฮเปอร์จีออเมตริก (Hypergeometric Distribution)

การทดลองแบบไฮเปอร์จีออเมตริก เกี่ยวข้องกับการทดลอง n ครั้ง ***คล้ายการทดลองแบบทวินาม และพหุนาม แต่การทดลองแต่ละครั้งไม่เป็นอิสระต่อกัน*** ความน่าจะเป็นของการเกิดเหตุการณ์เดียวกันในการทดลองแต่ละครั้งต่างกัน ตัวอย่างของการทดลองแบบไฮเปอร์จีออเมตริก ได้แก่การหยิบผลไม้จากถุงผลไม้ที่ประกอบด้วยส้ม 3 ลูก มะม่วง 2 ลูก และมังคุด 3 ลูก แบบไม่ใส่คืน

การทดลองแบบไฮเปอร์จีออเมตริก แบ่งออกเป็น 2 กรณีคือ (1) กรณีผลลัพธ์มี 2 แบบ (2) กรณีผลลัพธ์มี k แบบ

3.2.4.1 กรณีผลลัพธ์มี 2 แบบ (Bivariate Hypergeometric)

กรณีนี้จะคล้ายการแจกแจงแบบทวินาม ซึ่งผลลัพธ์มี 2 แบบ คือ "ความสำเร็จ" และ "ความล้มเหลว"

$$f(x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad \text{.........................(3.11)}$$

x = จำนวนครั้งที่ประสบความสำเร็จ

= 0, 1, 2,................., Min (k, n)

n = จำนวนครั้งของการทดลอง

N = จำนวนประชากรทั้งหมด

k = จำนวน "ความสำเร็จ" ใน N

N-k = จำนวน "ความล้มเหลว" ใน N

$$\mu = n\frac{k}{N} \qquad \text{........................(3.12)}$$

$$\sigma^2 = \frac{N-n}{N-1}\, n.\, \frac{k}{N}\left(1-\frac{k}{N}\right) \qquad \text{........................(3.13)}$$

ถ้า n << N มาก $\left(\frac{N-n}{N-1}\right) \sim 1$

ความน่าจะเป็นที่ผลลัพธ์คือ "ความสำเร็จ" (p) ในแต่ละครั้งของการทดลองจะมีค่าเกือบคงที่ กรณีนี้จะสามารถใช้การแจกแจงแบบทวินามประมาณค่าการแจกแจงแบบไฮเปอร์จีออเมตริกได้

พิสูจน์ $E(X) = \mu = \sum_{x=0}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$

$$= \sum_{x=0}^{n} x \frac{k!}{(k-x)!\,x!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= k \sum_{x=1}^{n} \frac{(k-1)!}{(k-x)!(x-1)!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= k \sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{(N-1)-(k-1)}{(n-1)-(x-1)}}{\frac{N}{n}\binom{N-1}{n-1}}$$

$$= k\frac{n}{N} \sum_{x=1}^{n} h(x-1 ; N-1, n-1, k-1)$$

ให้ $y = x-1$

ถ้า $x = 1 ;\ y = 0$

$x = n ; y = n-1$

$$E(X) = k\frac{n}{N} \sum_{y=0}^{n-1} h(y: N-1, n-1, k-1)$$

$$= n\frac{k}{N} \qquad \text{...................................(3.14)}$$

จาก $E(X^2) = E(X) + E(X(X-1))$

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1) \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=0}^{n} x(x-1) \frac{k!}{(k-x)!\,x!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= k(k-1) \sum_{x=0}^{n} \frac{(k-2)!}{(k-x)!(x-2)!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= k(k-1) \sum_{x=0}^{n} \binom{k-2}{x-2} \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$\binom{N}{n} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}$$

$$\binom{N-k}{n-x} = \binom{(N-2)-(k-2)}{(n-2)-(x-2)}$$

$$E(X(X-1)) = k(k-1)\frac{n(n-1)}{N(N-1)}\sum_{x=2}^{n}\binom{k-2}{x-2}\frac{\binom{(N-2)-(k-2)}{(n-2)-(x-2)}}{\binom{N-2}{n-2}}$$

$$= k(k-1)\frac{n(n-1)}{N(N-1)}\sum_{x=2}^{n} h(x-2\,;N-2,n-2,k-2)$$

ให้ $y = x-2$

$$E(X(X-1)) = k(k-1)\frac{n(n-1)}{N(N-1)}\sum_{y=0}^{n-2} h(y\,;N-2,n-2,k-2)$$

$$= k(k-1)\frac{n(n-1)}{N(N-1)}$$

$$\sigma^2 = E(X^2)-\mu^2$$

$$= \mu + k(k-1)\frac{n(n-1)}{N(N-1)} - \mu^2$$

$$= k\frac{n}{N} + k(k-1)\frac{n(n-1)}{N(N-1)} - k^2\frac{n^2}{N^2}$$

$$= \left(\frac{N-n}{N-1}\right) n.\frac{k}{N}\left(1-\frac{k}{N}\right) \qquad \text{........................(3.15)}$$

ตัวอย่างที่ 3.5 จากตัวอย่างที่ 3.3 ถ้าให้ X คือ จำนวนส้มที่หยิบได้จากการสุ่มเลือกผลไม้ แบบไม่ใส่คืน จงหา f(x), μ, σ^2

วิธีทำ

X = ตัวแปรสุ่มแบบไฮเปอร์จีออเมตริกแสดงจำนวนส้มที่หยิบได้

n = 4

N = 8

k = 3

x = 0, 1, 2, 3

$$f(x) = h(x\,;8,4,3) = \frac{\binom{3}{x}\binom{5}{4-x}}{\binom{8}{4}}$$

x	0	1	2	3	$\sum$
f(x)	$\frac{5}{70}$	$\frac{30}{70}$	$\frac{30}{70}$	$\frac{5}{70}$	1.0

$$\mu = n\frac{k}{N} = 4 \times \frac{3}{8} = \frac{3}{2}$$

$$\sigma^2 = \frac{N-n}{N-1}\, n.\, \frac{k}{N}\,(1-\frac{k}{N})$$

$$= \left(\frac{8-4}{8-1}\right) 4.\ \frac{3}{8}\,(1-\frac{3}{8})$$

$$= \frac{4}{7} \times 4 \times \frac{3}{8} \times \frac{5}{8} = \frac{15}{28}$$

ตัวอย่างที่ 3.6 บริษัทต้องการตรวจสอบสินค้าส่งออก 50 ชิ้น โดยการสุ่ม 5 ชิ้น และตั้งเกณฑ์ว่าถ้าพบสินค้าชำรุด 1 ชิ้น หรือมากกว่าผู้ซื้อจะส่งคืนสินค้าได้ บริษัทผู้ผลิตยอมรับว่ามีสินค้าชำรุด 20 % จงหาความน่าจะเป็นที่สินค้าจะถูกส่งคืน

วิธีทำ กรณีนี้เป็นการแจกแจงแบบไฮเปอร์จีออเมตริก

$N = 50$

$n = 5$

$k = \frac{20}{100} \times 50 = 10$ (จำนวนสินค้าชำรุดจากสินค้าทั้งหมด 50 ชิ้น)

$x \geq 1$

$P(X \geq 1) = 1-P(X = 0)$

$$P(X = 0) = \frac{\binom{10}{0}\binom{50-10}{5-0}}{\binom{50}{5}} = 0.3106$$

$P(X \geq 1) = 1-0.3106 = 0.6894$

ตัวอย่างที่ 3.7 ในการขนส่งยางรถยนต์ 5,000 เส้น ไปยังพ่อค้าขายส่ง มียาวชำรุด 1,000 เส้น ถ้าลูกค้าซื้อยาง 10 เส้น จากพ่อค้าขายส่ง จงหาความน่าจะเป็นที่จะพบยางชำรุด 3 เส้น

วิธีทำ

$N = 5{,}000$

$k = 1{,}000$

$n = 10$

$x = 3$

$$h(3\ ;\ 5{,}000,\ 10,\ 1{,}000) \quad = \quad \frac{\binom{1{,}000}{3}\binom{5{,}000-1{,}000}{10-3}}{\binom{5{,}000}{10}}$$

$$= \quad 0.2015$$

เนื่องจาก n << N มาก

สามารถใช้การแจกแจงแบบทวิมานประมาณค่าได้ดังนี้

$$p \quad = \quad \frac{k}{N} = \frac{1{,}000}{5{,}000} = 0.2$$

$$b(3\ ;\ 10,\ 0.2) \quad = \quad \frac{10\,!}{7\,!\,3\,!}(0.2)^3(0.8)^7$$

$$= \quad 0.2013$$

3.2.4.2 กรณีผลลัพธ์มี k แบบ (Multivariate Hypergeometric)

กรณีนี้จะคล้ายการแจกแจงแบบพหุนาม

ให้ $X_1, X_2, \ldots, X_k$ คือ จำนวนครั้งที่ผลลัพธ์เป็น $E_1, E_2, \ldots, E_k$ ตามลำดับในการทดลอง n ครั้ง และ $a_1, a_2, \ldots, a_k$ คือ จำนวนสมาชิกของประชากรใน $E_1, E_2, \ldots, E_k$ ตามลำดับ

Outcomes	No. of Success ใน N	No. of Success ใน n
E_1	a_1	x_1
E_2	a_2	x_2
.	.	.
.	.	.
E_k	a_k	x_k
$\sum$	N	n

$f(x_1, x_2, \ldots, x_k\ ;\ a_1, a_2, \ldots, a_k, N, n)$

$$= \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\cdots\binom{a_k}{x_k}}{\binom{N}{n}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.16)$$

เมื่อ $x = 0, 1, \ldots, \mathrm{Min}\,(a_i, n)$ เมื่อ $i = 1, 2, \ldots, k$

$$= \sum_{i=1}^{k} a_i = N$$

$$= \sum_{i=1}^{k} x_i = n$$

μ และ σ^2 มีค่าดังตาราง

ตัวแปรสุ่ม	μ	σ^2
X_1	$n.\frac{a_1}{N}$	$\left(\frac{N-n}{N-1}\right)n.\frac{a_1}{N}\left(1-\frac{a_1}{N}\right)$
X_2	$n.\frac{a_2}{N}$	$\left(\frac{N-n}{N-1}\right)n.\frac{a_2}{N}\left(1-\frac{a_2}{N}\right)$
.	.	.
.	.	.
X_k	$n.\frac{a_k}{N}$	$\left(\frac{N-n}{N-1}\right)n.\frac{a_k}{N}\left(1-\frac{a_k}{N}\right)$

ตัวอย่างที่ 3.8 กลุ่มคน 10 คน ประกอบด้วยคนเลือดกรุ๊ป O 3 คน A 4 คน และ B 3 คน ถ้าสุ่มตรวจเลือด 5 คน พบว่า x มีค่าดังตาราง จงหาความน่าจะเป็นที่จะสุ่มได้ผลลัพธ์ดังกล่าว

กลุ่มเลือด	a_i	x_i
O	3	1
A	4	2
B	3	2
$\sum$	N = 10	n = 5

f (1, 2, 2 ; 3, 4, 3, 10, 5)

$$= \frac{\binom{3}{1}\binom{4}{2}\binom{3}{2}}{\binom{10}{5}} = \frac{3}{14}$$

3.2.5 การแจกแจงแบบปัวซอง (Poisson Distribution)

ปัวซองใช้อธิบายการแจกแจงจำนวนครั้งของความสำเร็จในช่วงเวลาหนึ่ง หรือต่อหนึ่งหน่วยพื้นที่ เช่น จำนวนคนที่เข้าห้องสมุดใน 1 วัน หรือจำนวนรถแทรกเตอร์ต่อพื้นที่ 1 ไร่ เป็นต้น

คุณสมบัติที่สำคัญของตัวแปรสุ่มปัวซองคือ จำนวนครั้งของความสำเร็จในช่วงเวลาหนึ่งเป็นอิสระต่อจำนวนครั้งของความสำเร็จในช่วงเวลาอื่น

$$f(x) = p(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!} \quad \text{เมื่อ } x = 0, 1, 2 \ldots \ldots \ldots \ldots (3.17)$$

$$F(r) = P(X \le r) = P(r;\mu)$$

$$= \sum_{x=0}^{r} p(r;\mu) \quad \ldots \ldots \ldots \ldots (3.18)$$

ค่า cdf ของปัวซองแสดงอยู่ในตารางที่ A2 ของภาคผนวก

$$E(X) = \mu \quad \ldots \ldots \ldots \ldots (3.19)$$

$$\sigma^2 = \mu \quad \ldots \ldots \ldots \ldots (3.20)$$

พิสูจน์

$$E(X) = \sum_{x=0}^{\infty} x\frac{e^{-\mu}\mu^{x}}{x!}$$

$$= \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^{x}}{(x-1)!}$$

$$= \mu\sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^{x-1}}{(x-1)!}$$

ให้ $y = x-1$

$$E(X) = \mu\sum_{y=0}^{\infty} \frac{e^{-\mu}\mu^{y}}{y!} = \mu$$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\mu}\mu^{x}}{x!}$$

$$= \mu^2\sum_{x=2}^{\infty} \frac{e^{-\mu}\mu^{x-2}}{(x-2)!}$$

ให้ $y = x-2$

$$E(X(X-1)) = \mu^2\sum_{y=0}^{\infty} \frac{e^{-\mu}\mu^{y}}{y!}$$

$$= \quad \mu^2$$

$$\sigma^2 = \mu + E(X(X-1)) - \mu^2$$

$$= \mu + \mu^2 - \mu^2 = \mu$$

ถ้า $n \rightarrow \infty$ และ $p \rightarrow 0$ หรือ $p \rightarrow 1$

และ $\mu = np$ มีค่าคงที่

สามารถใช้ $p(x;\mu)$ เป็นตัวประมาณค่าของ $b(x;n,p)$ ได้

พิสูจน์ $\mu = np$

$$p = \frac{\mu}{n}$$

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

$$= \frac{n(n-1)........(n-x+1)}{x!} \left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x}$$

$$= \frac{n(n-1)........(n-x+1)}{n^x} \left(\frac{\mu^x}{x!}\right) \left(1-\frac{\mu}{n}\right)^n \left(1-\frac{\mu}{n}\right)^{-x}$$

ถ้า $n \rightarrow \infty$

$$\lim_{n\rightarrow\infty} \frac{n(n-1)........(n-x+1)}{n^x}$$

$$= \lim_{n\rightarrow\infty} 1(1-1\,(1-\frac{1}{n}).........(1-\frac{x-1}{n}) = 1$$

$$\lim_{n\rightarrow\infty} \left(1-\frac{\mu}{n}\right)^{-x} = 1$$

จากนิยามของ e

$$\lim_{n\rightarrow\infty} \left(1-\frac{\mu}{n}\right)^n = \lim_{n\rightarrow\infty} \{[1+\frac{1}{-\frac{n}{\mu}}]^{-\frac{n}{\mu}}\}^{-\mu}$$

$$= e^{-\mu}$$

$$b(x;n,p) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{เมื่อ } x = 0, 1, 2, ... (3.20)$$

ตัวอย่างที่ 3.9 โดยเฉลี่ยมีเรือบรรทุกน้ำมันเล่นเข้าสู่ท่าเรือแห่งหนึ่ง 10 ลำต่อวัน ถ้าอุปกรณ์ขนถ่ายน้ำมันที่ท่าเรือสามารถขนถ่ายน้ำมันได้ 15 ลำต่อวัน จงหาความน่าจะเป็นที่ท่าเรือจะขนถ่ายน้ำมันไม่ทัน

วิธีทำ

X = จำนวนเรือบรรทุกน้ำมันที่แล่นเข้าท่าเรือต่อวัน

μ = 10 ลำต่อวัน

$$P(X > 15) = 1 - P(X \le 15)$$

$$= 1 - \sum_{x=0}^{15} p(x; 10)$$

$$= 1 - 0.9513$$

$$= 0.0487$$

ตัวอย่างที่ 3.10 โรงงานผลิตแก้ว ผลิตแก้วเสีย 1 ใบ ในทุก ๆ 1,000 ใบ จงหาความน่าจะเป็นที่โรงงานจะผลิตแก้วเสียน้อยกว่า 7 ใบใน 8,000 ใบ

วิธีทำ X = จำนวนแก้วเสียในแก้วที่ผลิต 8,000 ใบ

$$n = 8{,}000 \qquad (n \rightarrow \infty)$$

$$p = \frac{1}{1{,}000} \qquad (p \rightarrow \infty)$$

$$P(X < 7) = \sum_{x=0}^{6} b\left(x\,;\, 8{,}000,\, \frac{1}{1{,}000}\right)$$

= [คำนวณยาก]

ประมาณโดยใช้ปัวซอง

$$\mu = np = 8{,}000 \times \frac{1}{1{,}000} = 8$$

$$P(X < 7) = \sum_{x=0}^{6} p(x\,;\, 8)$$

$$= 0.313$$

3.2.6 การแจกแจงแบบทวินามลบและแบบเรขาคณิต (Negative Binomial and Geometric Distributions)

3.2.6.1 การแจกแจงแบบทวินามลบ

การแจกแจงแบบทวินามลบ นิยามค่าตัวแปรสุ่ม X ตรงกันข้ามกับการแจกแจงแบบทวินาม โดยมีคุณสมบัติที่สำคัญซึ่งแตกต่างกันดังตาราง

	Binomial	Negative Binomial
ตัวแปรสุ่ม	X (No. of success in n trials)	X (No. of trial with k successes)
ค่าที่กำหนดให้	n, p	k, p
ค่าของ x	0, 1, 2,........, n	k, k+1,..........

$$f(x) = b^*(x\,;k,p) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \quad \text{เมื่อ } x = k, k+1, \qquad \ldots\ldots\ldots\ldots(3.21)$$

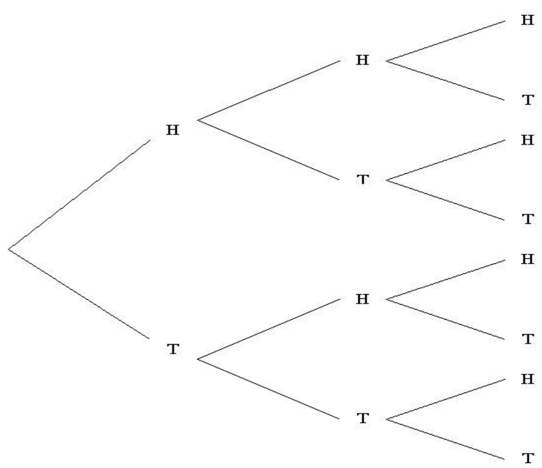
เนื่องทวินามลบ กำหนดว่า ความสำเร็จครั้งที่ k เกิดในการทดลองครั้งที่ x จึงทำให้คิดจำนวนวิธี (Combination) เพียง $\binom{x-1}{k-1}$ เท่านั้น

$$\mu = \frac{k}{p}$$

$$\sigma^2 = \frac{k(1-p)}{p^2}$$

ตัวอย่างที่ 3.11 ในการโยนเหรียญ 3 อัน 5 ครั้ง จงหาความน่าจะเป็นที่ออกหัว 3 อัน ครั้งที่ 2

Sample Space {S} ของ โยนเหรียญ 3 อัน แต่ละครั้ง



{Sample Space} = HHH, HHT, HTH, HTT
THH, THT, TTH, TTT

ถ้าเหตุการณ์ที่สนใจคือ HHH

{Sample Space} = {HHH, HHH'}

$$p = P(S) = P(HHH) = \left(\frac{1}{2}\right)^3 = \frac{1}{8} = \text{Prob. of Success}$$

$$q = P(F) = P(HHH') = \frac{7}{8} = \text{Prob. of Failure}$$

จำนวนเหตุการณ์โยนเหรียญ 3 อัน 5 ครั้ง $= 2^5 = 32$

SSSSS
SSSSF
SSSFS
SSSFF
SSFSS
SSFSF
SSFFS
SSFFF++
SFSSS
SFSSF
SFSFS
SFSFF++
SFFSS
SFFSF
SFFFS++*
SFFFF++
________(16 เหตุการณ์)

FSSSS
FSSSF
FSSFS
FSSFF++
FSFSS
FSFSF++
FSFFS++*
FSFFF
FFSSS
FFSSF++
FFSFS++*
FFSFF
FFFSS++*
FFFSF
FFFFS
FFFFF
________(16 เหตุการณ์)

* คือ Sample Points ของทวินามลบ

++ คือ Sample Points ของทวินาม

$$b(x\,;n,p) = b(2;5,\tfrac{1}{8}) = \binom{5}{2}\left(\frac{1}{8}\right)^2\left(\frac{7}{8}\right)^3$$

$$= \frac{5!}{2!3!}\left(\frac{1}{8}\right)^2\left(\frac{7}{8}\right)^3 = 10\left(\frac{1}{8}\right)^2\left(\frac{7}{8}\right)^3 = 0.1047$$

$$b^*(x\,;k,p) = b^*(5;2,\tfrac{1}{8}) = \binom{5-1}{2-1}\left(\frac{1}{8}\right)^2\left(\frac{7}{8}\right)^3$$

$$= \frac{4!}{1!3!}\left(\frac{1}{8}\right)^2\left(\frac{7}{8}\right)^3 = 4\left(\frac{1}{8}\right)^2\left(\frac{7}{8}\right)^3 = 0.0419$$

3.2.6.2 การแจกแจงแบบเรขาคณิต

การแจกแจงแบบเรขาคณิต คือ ทวินามลบกรณีที่ k=1

$$
\begin{aligned}
f(x) &= b^*(x\,;1,p) \\
&= \binom{x-1}{1-1} p^{1}\ (1-p)^{x-1} \\
&= p\,(1-p)^{x-1} \\
&= g\,(x\,;p) \text{ เมื่อ } x = 1, 2, \quad \text{...............................(3.22)} \\
\mu &= \frac{1}{p} \\
\sigma^2 &= \frac{(1-p)}{p^2}
\end{aligned}
$$

ตัวอย่างที่ 3.11 ในการผลิต พบว่ามีผลิตภัณฑ์เสีย 1 ชิ้น ในทุก ๆ 100 ชิ้น จงหาความน่าจะเป็น ในการตรวจสอบ 5 ชิ้น พบผลิตภัณฑ์เสียชิ้นแรก

วิธีทำ

$$
\begin{aligned}
x = 5\,;p &= \frac{1}{100} \\
g\,(x\,;p) &= g\,(5\,;\frac{1}{100}) \\
&= (\frac{1}{100})\ (1-\frac{1}{100})^4 \\
&= 0.096
\end{aligned}
$$

3.2.7 สรุปคุณสมบัติที่สำคัญของการแจกแจงตัวแปรสุ่มไม่ต่อเนื่องแบบต่าง ๆ

คุณสมบัติที่สำคัญของการแจกแจงตัวแปรสุ่มไม่ต่อเนื่องแบบต่าง ๆ แสดงอยู่ในตาราง ที่ 3.1

ตารางที่ 3.1 ตารางสรุปคุณสมบัติที่สำคัญของการแจกแจงตัวแปรสุ่มไม่ต่อเนื่องแบบต่าง ๆ

แบบที่	การแจกแจง	ค่าของ X	f(x)	μ	σ^2
1	แบบเอกรูป หรือ แบบสม่ำเสมอ	$(x_1, x_2, ..., x_k)$	$u(x ; k)= \frac{1}{k}$	$\frac{1}{k}\sum_{i=1}^{k} x_i$	$\frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu^2$
2	แบบทวินาม	$0, 1, 2, ..., n$	$b(x ; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$	np	npq
3	แบบพหุนาม	$x_1 = 0, 1, 2, ..., n$ $x_2 = 0, 1, 2, ..., n$ . ... $x_k = 0, 1, 2, ..., n$	$f(x_1, x_2, ..., x_k; p_1, p_2, ..., p_k, n)$ $= \frac{n!}{x_1! x_2! ... x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$ เมื่อ $x_1+x_2+.....+x_k = n$ $p_1+p_2+.....+p_k = 1.0$	np_1 np_2 . np_k	np_1q_1 np_2q_2 . np_kq_k
4	แบบไฮเปอร์จีออเมตริก กรณีผลลัพธ์มี 2 แบบ	0,1,2,..,Min(k,n)	$h(x ; N, n, k)$ $= \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$	$n\frac{k}{N}$	$\left(\frac{N-n}{N-1}\right) n \cdot \frac{k}{N}(1-\frac{k}{N})$
5	แบบไฮเปอร์จีออเมตริก กรณีผลลัพธ์ k แบบ	$x_1 = 0,1,2,......,$ $Min(a_1,n)$ $x_2 = 0,1,2,......,$ $Min(a_2,n)$. $x_k = 0,1,2,......,$ $Min(a_k,n)$	$f(x_1, x_2, ..., x_k ; a_1, a_2, ..., a_k, N, n)$ $= \frac{\binom{a_1}{x_1}\binom{a_2}{x_2} ...... \binom{a_k}{x_k}}{\binom{N}{n}}$ เมื่อ $x_1+x_2+.....+x_k = n$ $a_1+a_2+.....+a_k = N$	$n\frac{a_1}{N}$ $n\frac{a_2}{N}$. $n\frac{a_k}{N}$	$\left(\frac{N-n}{N-1}\right) n \cdot \frac{a_1}{N}(1-\frac{a_1}{N})$ $\left(\frac{N-n}{N-1}\right) n \cdot \frac{a_2}{N}(1-\frac{a_2}{N})$. $\left(\frac{N-n}{N-1}\right) n \cdot \frac{a_k}{N}(1-\frac{a_k}{N})$
6	แบบปัวซอง	0,1,2,...............	$p(x ; \mu) = \frac{e^{-\mu}\mu^x}{x!}$	μ	μ
7	แบบทวินามลบ	k,k+1,.............	$b^*(x ; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
8	แบบเรขาคณิต	1,2,.................	$g(x ; p) = p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

3.3 ตัวแปรสุ่มต่อเนื่อง

ตัวแปรสุ่มต่อเนื่องได้แก่ตัวแปรสุ่มที่แสดงค่าความสูง น้ำหนัก อุณหภูมิ ระยะทางหรืออายุการใช้งานของอุปกรณ์ ซึ่งมีค่าต่อเนื่อง สิ่งที่จำเป็นต้องรู้เกี่ยวกับตัวแปรสุ่มต่อเนื่องคือ

- ตัวแปรสุ่ม X มีค่าต่ำสุด-สูงสุดเท่าไร
- f(x) และ F(x) มีลักษณะการแจกแจงอย่างไร มีพารามิเตอร์อะไรบ้าง และหาพารามิเตอร์ได้อย่างไร

  $f(x) = \frac{dF}{dx}(x)$ (Probability Density Function, pdf)

  $F(x) = \int_{-\infty}^{x} f(t)\, dt$ (Cumulative Distribution Function, cdf)
- พารามิเตอร์ของ X เช่น μ และ σ^2 มีค่าเท่าไร

เมื่อทราบคุณสมบัติเกี่ยวกับการแจกแจงของตัวแปรสุ่มต่อเนื่องแล้ว จะสามารถนำไปวิเคราะห์เพื่อตอบคำถามเกี่ยวกับความน่าจะเป็นที่ต้องการได้ เช่น

- ถ้ากำหนดค่า x ให้ จงหา $P(X \le x)$ หรือ $P(X > x)$
- ถ้ากำหนดค่า $P(X \le x) = F(x)$ จงหาค่า x หรือหา Inverse Function ของ x, $F^{-1}(x)$

การแจกแจงของตัวแปรสุ่มต่อเนื่องที่จะกล่าวถึงในบทนี้มี 6 ชนิดคือ

(1) การแจกแจงแบบเอกรูป หรือแบบสม่ำเสมอ (Uniform Distribution)

(2) การแจกแจงแบบปกติ (Normal Distribution)

(3) การแจกแจงแบบแกมม่า (Gamma Distribution)

(4) การแจกแจงแบบเอกซ์โปเนนเชียล (Exponential Distribution)

(5) การแจกแจงแบบไคสแควร์ (Chi-Squared Distribution, χ^2)

(6) การแจกแจงแบบไวบูล (Weibull Distribution)

การแจกแจงแบบเอกซ์โปเนนเชียลและไคสแควร์ เป็นกรณีพิเศษของการแจกแจงแบบแกมม่า ส่วนการแจกแจงแบบไวบูลจะคล้ายกับแกมม่าและเอกซ์โปเนนเชียล

- ถ้า X คือตัวแปรสุ่มต่อเนื่อง สิ่งสำคัญของตัวแปรสุ่มต่อเนื่องที่แตกต่างจากตัวแปรสุ่มไม่ต่อเนื่องคือ $P(X = 100) = 0$

  ดังนั้น $P(X \le 100) = P(X < 100)$

3.3.1 การแจกแจงแบบเอกรูปหรือแบบสม่ำเสมอ (Uniform Distribution)

การแจกแจงตัวแปรสุ่มต่อเนื่องแบบเอกรูปหรือสม่ำเสมอ จะเกี่ยวข้องกับการทดลองซึ่งโอกาสความน่าจะเป็นที่จะเกิดผลลัพธ์ใด ๆ มีค่าเท่ากัน ***แต่ปกติทั่วไปตัวแปรสุ่มต่อเนื่องมักมีการแจกแจงแบบไม่สม่ำเสมอ*** เช่น

- อุณหภูมิเฉลี่ยประจำวันของอากาศที่กำแพงแสนมีค่าอยู่ระหว่าง 15-40°C แต่ความน่าจะเป็นที่อุณหภูมิอยู่ระหว่าง 25-30°C จะมากกว่าความน่าจะเป็นที่อุณหภูมิอยู่ระหว่าง 15-20°C หรือ 35-40°C แสดงว่าอุณหภูมิเฉลี่ยประจำวันแจกแจงแบบไม่สม่ำเสมอ
- นิสิตชายที่วิทยาเขตกำแพงแสนมีความสูงระหว่าง 145-185 ม. แต่ความน่าจะเป็นที่ความสูงนิสิตชายอยู่ระหว่าง 165-175 ม. จะมากกว่าความน่าจะเป็นที่ความสูงนิสิตชายอยู่ระหว่าง 145-155 ม. หรือ 175-185 ม. แสดงว่าความสูงนิสิตชายของวิทยาเขตกำแพงแสนมีการแจกแจงแบบไม่สม่ำเสมอ

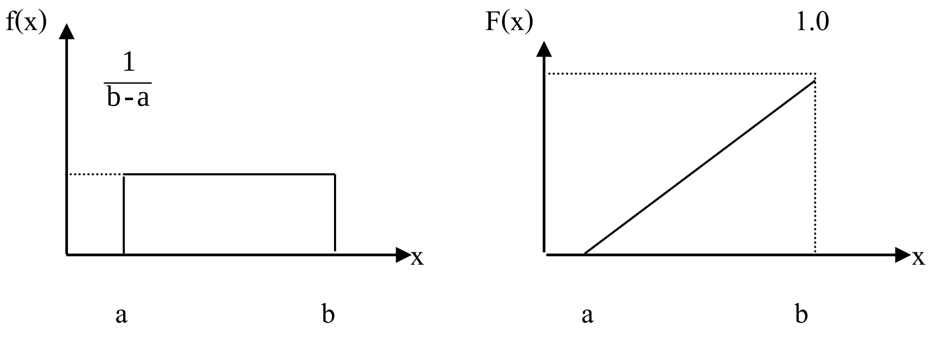
ให้ X = ตัวแปรสุ่มต่อเนื่องแบบเอกรูป

$$f(x) = \frac{1}{(b-a)} \quad \text{เมื่อ } a \le x < b \quad \text{.......................................(3.23)}$$

$$F(x) = \int_a^x \frac{1}{(b-a)}\,dx$$

$$= \frac{x-a}{b-a} \quad \text{เมื่อ } a \le x < b \quad \text{.......................................(3.24)}$$

กราฟ f(x) และ F(x) แสดงอยู่ในรูปที่ 3.2



รูปที่ 3. 2 กราฟ f(x) และ F(x) ของการแจกแจงแบบเอกรูป

$$\mu = E(X) = \int_a^b x.\, f(x)\,dx = \frac{b^2 - a^2}{2(b-a)} = \frac{1}{2}(a+b) \quad \text{..............(3.25)}$$

$$\sigma^2 = E(X^2) - \mu^2 = \int_a^b x^2 . f(x)\,dx - \frac{1}{4}(a+b)^2$$
$$= \frac{b^3 - a^3}{3(b-a)} - \frac{1}{4}(a+b)^2$$
$$= \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4}$$
$$= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12} \qquad \text{.........................(3.26)}$$

ตัวอย่างที่ 3.12 ความดันภายในท่อก๊าซจะเปลี่ยนแปลงอยู่ในช่วง 200-300 psi ถ้าความน่าจะเป็นของความดันที่ระดับใด ๆ มีค่าคงที่ และวิศวกรกำหนดว่าความดันในท่อจะอยู่ในเกณฑ์ปลอดภัยเมื่อมีค่าไม่เกิน 280 psi จงหา

(ก) ความน่าจะเป็นที่ความดันภายในท่อจะอยู่ในเกณฑ์ปลอดภัย

(ข) ความเสี่ยงที่ความดันภายในท่อจะไม่ปลอดภัย

(ค) μ และ σ

(ง) ความน่าจะเป็นที่ความดันจะอยู่ในช่วง $\mu \pm \sigma$

(จ) จงหาความดัน ถ้าความเสี่ยงที่ความดันภายในท่อจะไม่ปลอดภัยเท่ากับ 0.1

วิธีทำ (ก) X = ความดันภายในท่อก๊าซ, psi

$$f(x) = \frac{1}{300-200} = \frac{1}{100} \qquad \text{เมื่อ } 200 \le x < 300$$

$$P(X \le 280) = \int_{200}^{280} \frac{1}{100}\,dx$$
$$= \frac{280-200}{100} = 0.8$$ (ความน่าจะเป็นที่ความดันอยู่ในเกณฑ์ปลอดภัย

(ข) $P(X>280) = 1 - P(X \le 280)$ (ความเสี่ยงที่ความดันจะไม่ปลอดภัย)

$= 1 - 0.8 = 0.2$

(ค) $\mu = \frac{1}{2}(a+b) = \frac{1}{2}(200+300)$

$= 250$ psi

$\sigma^2 = \frac{1}{12}(a-b)^2 = \frac{1}{12}(300-200)^2$

$= \frac{1,000}{12} = 833 \text{ psi}^2$

$\sigma = 28.9$ psi

(ง) $P(\mu - \sigma \le x < \mu + \sigma)$

$= P(250-28.9 \le X < 250+28.9)$

$$= P(221.1 \le X < 278.9)$$

$$= \int_{221.1}^{278.9} \frac{1}{100}\,dx$$

$$= \frac{278.9-221.1}{100} = 0.578$$

(จ) $P(X > x) = 0.1$

$$P(X \le x) = 0.9 = \int_{200}^{x} \frac{1}{100}\,dx$$

$$\frac{x-200}{100} = 0.9$$

$$x = 0.9 \times 100 + 200 = 290 \text{ psi}$$

3.3.2 การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงแบบปกติ ถือเป็นการแจกแจงที่มีความสำคัญและใช้มากที่สุดในวิชาสถิติ ลักษณะสำคัญของการแจกแจงแบบปกติคือ โค้งจะสมมาตร (Symmetry) รอบจุดศูนย์กลางดังแสดงในรูปที่ 5.2 การแจกแจงแบบปกติมีชื่อเรียกอีกอย่างหนึ่งว่า Gaussian Distribution

$$f(x) = n(x\,;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty \le x < \infty \quad ....(3.27)$$

μ = พารามิเตอร์ที่แสดงตำแหน่งจุดศูนย์กลาง (Location Parameter) = E(X)

σ^2 = พารามิเตอร์ที่แสดงมาตราส่วน (Scale Parameter)

หรือการกระจาย = $E(X^2) - \mu^2$

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx \quad ....(3.28)$$

$$P(x_1 \le X < x_2) = F(x_2)-F(x_1) = \int_{x_1}^{x_2} f(x)\,dx = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx \quad ....(3.29)$$

3.3.2.1 พื้นที่ใต้โค้งปกติ

จากสมการที่ 5.6 การอินทีเกรท (Integrate) หาค่า F(x) หรือการหา Inverse Function, $F^{-1}(x)$, ทำได้ค่อนข้างยากถ้าไม่มีเครื่องคำนวณช่วย ดังนั้นเพื่อช่วยให้การคำนวณต่าง ๆ เกี่ยวกับการแจกแจงปกติทำได้ง่ายขึ้นจะแปลงตัวแปรสุ่มปกติ (X) เป็นตัวแปรสุ่มปกติมาตราฐาน (Standard Normal Variate) Z ซึ่งมีวิธีการดังนี้

ถ้า X เป็นตัวแปรสุ่มปกติมี μ, σ^2 และ f(x) ดังสมการที่ 5.6

$$\text{ให้ } Z = \frac{X-\mu}{\sigma} \quad ....(3.30)$$

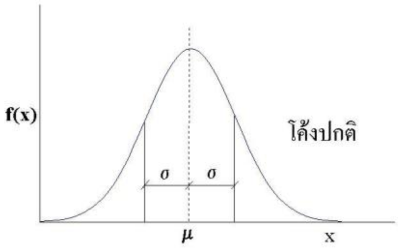
Z = ตัวแปรสุ่มปกติมาตราฐานมี $\mu_Z = 0$ และ $\sigma_Z = 1$

พิสูจน์ $E(Z) = \dfrac{E(X)-\mu}{\sigma}$

$\mu_z = 0$

$Var(Z) = \dfrac{Var(X)}{\sigma^2}$

$\sigma_z^2 = 1$

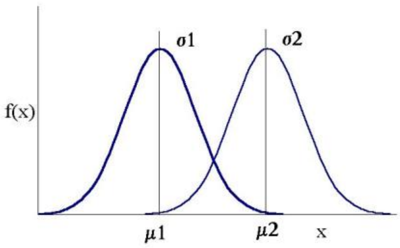


คุณสมบัติของโค้งปกติ

1. จุดที่โค้งสูงสุด (Mode) = μ
2. Symmetry รอบ μ
3. ที่ $x = \mu \pm \sigma$ คือจุดเปลี่ยนโค้ง
4. พ.ท.ใต้โค้ง = 1.00

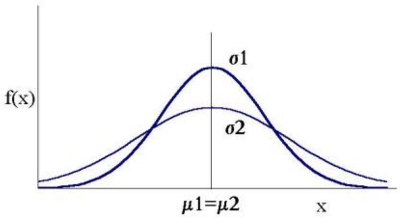
รูปที่ 3.3 โค้งปกติ

โค้งปกติกรณีที่ μ และ σ ต่างกัน แสดงอยู่ในรูปที่ 5.3



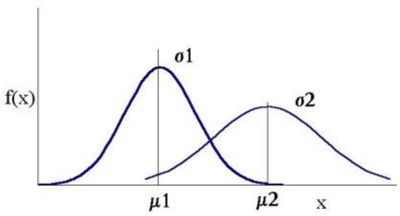
(1) $\mu_1 < \mu_2$

$\sigma_1 = \sigma_2$



(2) $\mu_1 = \mu_2$

$\sigma_1 < \sigma_2$



(3) $\mu_1 < \mu_2$

$\sigma_1 < \sigma_2$

รูปที่ 3.4 โค้งปกติกรณีที่ μ และ σ ต่างกัน

การหา f(z)

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

$$dz = \frac{1}{\sigma}dx$$

$$\frac{dx}{dz} = \sigma$$

จากรูปที่ 3.3 ที่ dx ใด ๆ

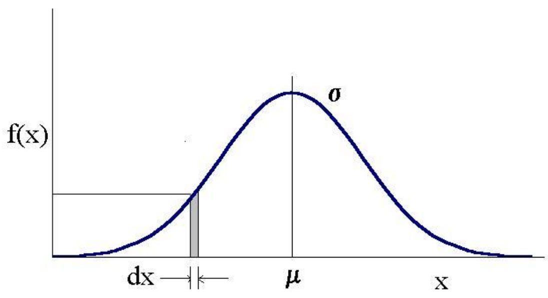
$$f(x)\,dx = f(z)\,dz$$

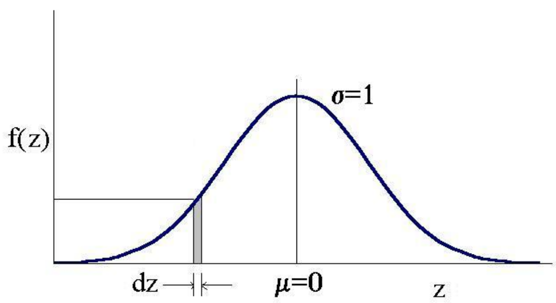
$$f(z) = f(x)\frac{dx}{dz}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \quad \text{เมื่อ } -\infty \leq z < \infty$$

$$= n\,(Z\,;\,0,\,1)$$





รูปที่ 3.5 กราฟเปรียบเทียบ f(x) และ f(z)

$$P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = \int_{z_1}^{z_2} f(z)\,dz = P(z_1 \leq Z < z_2) \quad ....(3.31)$$

เมื่อ $z_1 = \dfrac{x_1 - \mu}{\sigma}$

$z_2 = \dfrac{x_2 - \mu}{\sigma}$

ตาราง $F(z) = P(Z \leq z)$ เมื่อ $-3.4 \leq Z < 3.4$ แสดงอยู่ในตารางที่ A3

ตัวอย่างที่ 5.13 ให้ X = ตัวแปรสุ่มปกติ

มี $\mu = 18$

$\sigma = 2.5$

จงหา (ก) $P(X \leq 15)$

(ข) $P(X > 15)$, $P(X \leq -15)$, $P(X > -15)$

(ค) ค่าของ k ถ้า $P(X \leq k) = 0.2578$

(ง) $P(17 \leq X < 21)$

(จ) ค่าของ k ถ้า $P(X > k) = 0.1539$

วิธีทำ

(ก) $P(X \leq 15) = P(z \leq \frac{15-18}{2.5})$

$= P(z \leq -1.2) = F(-1.2)$

$= 0.1151$

(ข) $P(X > 15) = 1-P(X \leq 15)$

$= 1-0.1151$

$= 0.8849$

$P(X \leq -15) = P(X > 15)$

$= 0.8849$

$P(X > -15) = P(X \leq 15)$

$= 0.1151$

(ค) $P(X \leq k) = 0.2578 = F(z)$

$= P(Z \leq -0.65)$

$z = -0.65$

$\frac{x-18}{2.5} = -0.65$

$x = -0.65 \times 2.5 + 18$

$= 16.375 = k$

(ง) $P(17 \le X < 21) = P\left(\frac{17-18}{2.5} \le Z < \frac{21-18}{2.5}\right)$

$= P(-0.4 \le Z < 1.2)$

$= F(1.2) - F(-0.4)$

$= 0.8849 - 0.3446$

$= 0.5403$

(จ) $P(X > k) = 0.1539$

$P(X \le k) = 1-0.1539$

$= 0.8461$

$= P(Z \le 1.02)$

$z = \frac{x-18}{2.5} = 1.02$

$x = 1.02 \times 2.5 + 18$

$= 20.55 = k$

ทฤษฎี Chebyshev

ถ้าการแจกแจงเป็นแบบปกติ $P(\mu-2\sigma \le X < \mu+2\sigma) \ge 1-\frac{1}{4} \ge 0.75$

$P(\mu-2\sigma \le X < \mu+2\sigma) = P(18 - 2 \times 2.5 \le X < 18 + 2 \times 2.5)$

$= P(13 \le X < 23)$

$= P\left(\frac{13-18}{2.5} \le Z < \frac{23-18}{2.5}\right)$

$= P(-2 \le Z < 2)$

$= 0.9772 - 0.0228$

$= 0.9544 > 0.75$

แสดงว่าทฤษฎี Chebyshev เป็นจริง

ตัวอย่างที่ 3.14 ถ้าคะแนนวิชาสถิติมีการแจกแจงปกติโดยมี $\mu = 74$; $\sigma = 7.9$

จงหา (ก) คะแนนต่ำสุดที่จะสอบผ่าน ถ้าความน่าจะเป็นที่จะได้ F เท่ากับ 10%

(ข) คะแนนสูงสุดที่จะได้ B ถ้าความน่าจะเป็นที่จะได้ A เท่ากับ 5%

วิธีทำ ให้ X = ตัวแปรสุ่มแสดงคะแนนวิชาสถิติ

(ก) ให้คะแนนต่ำสุดที่สอบได้ผ่าน = x_1

ถ้าต่ำสุด 10% ได้ F

$P(F) = P(X < x_1) = 0.1$

$= P(Z < -1.28)$

$z_1 = -1.28$

$\dfrac{x_1 - 74}{7.9} = -1.28$

$x_1 = 63.88$ คะแนน

(ข) คะแนนสูงสุดที่จะได้ $B = x_2$

$P(X > x_2) = 0.05$

$P(X \le x_2) = 1-0.05 = 0.95$

$= P(Z \le 1.645)$

$z_2 = 1.645$

$\dfrac{x_2 - 74}{7.9} = 1.645$

$x_2 = 1.645 * 7.9 + 74$

$= 86.99$ คะแนน

ตัวอย่างที่ 3.15 ตู้กดน้ำอัดลมตั้ง เครื่องรินน้ำอัดลมแบบอัตโนมัติ ให้ใส่น้ำถ้วยละ 7 ออนซ์โดยเฉลี่ยโดยมี $\sigma = 0.5$ ออนซ์

จงหา (ก) ความน่าจะเป็นที่น้ำอัดลมในถ้วยจะมีค่าอยู่ระหว่าง 6.7-7.3 ออนซ์

(ค) ความน่าจะเป็นที่เครื่องจะรินล้นถ้วย ถ้าใช้ถ้วยขนาด 8 ออนซ์

วิธีทำ

(ก) $P(6.7 \le X < 7.3$ ออนซ์$)$

$= P\left(\dfrac{6.7-7}{0.5} \le Z < \dfrac{7.3-7}{0.5}\right)$

$= P(-0.6 \le Z < 0.6)$

$= 0.7257 - 0.2743 = 0.4514$

(ข) $P(X > 8) = P\left(Z > \dfrac{8-7}{0.5}\right)$

$= P(Z > 2)$

$= 1-P(Z \le 2)$

$= 1-0.9772$

$\therefore$ ความน่าจะเป็นที่เครื่องจะรินล้นถ้วยขนาด 8 ออนซ์

$= 0.0228$

3.3.2.2 การใช้การแจกแจงแบบปกติประมาณค่าความน่าจะเป็นของการแจกแจงแบบทวินาม (Normal Approximation to Binomial)

ให้ X = ตัวแปรสุ่มแบบทวินามซึ่งมี μ = np ; σ^2 = npq

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad \text{เมื่อ } x = 0, 1, 2, \ldots\ldots\ldots n$$

ตามที่กล่าวมาแล้วในหัวข้อที่ 4.6 (การแจกแจงแบบปัวซอง)

ถ้า $n \rightarrow \infty$

$p \rightarrow 0$ หรือ 1

$$b(x; n, p) \approx p(x; \mu) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{เมื่อ } \mu = np$$

การแจกแจงแบบปกติสามารถใช้ประมาณค่าการแจกแจงแบบทวินามได้

เมื่อ $p \not\rightarrow 0$ หรือ $p \not\rightarrow 1$ และจะประมาณค่าได้ดีเมื่อ $n \rightarrow \infty$

หรือ n เล็ก $p \rightarrow \frac{1}{2}$

หรือ np หรือ nq $>$ 5

สามารถใช้การแจกแจงแบบปกติประมาณค่าความน่าจะเป็นของตัวแปรสุ่มแบบทวินาม X ได้โดยแปลง X เป็น Z ดังนี้

ให้ $Z = \frac{X - np}{\sqrt{npq}} \rightarrow n(z; 0, 1)$

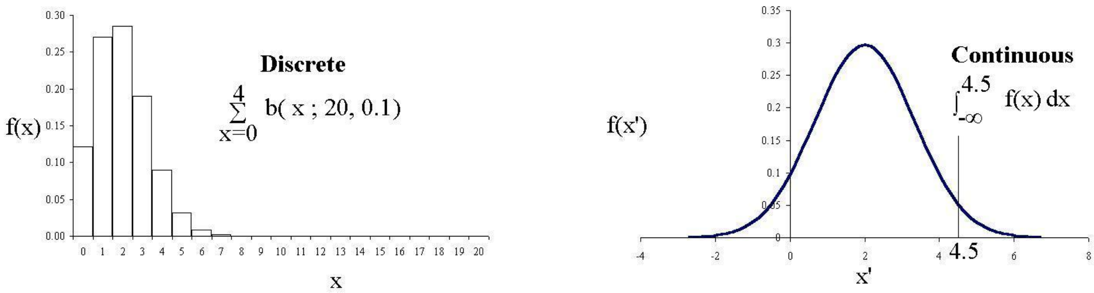
ตัวอย่างที่ 3.16 ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องแบบทวินามมี n = 20, p = 0.1 จงหาความผิดพลาดในการประมาณค่า $P(X \leq 4)$ หรือ $\sum_{x=0}^{4} b(x;20, 0.1)$ โดยใช้การแจกแจงแบบปกติ

วิธีทำ

$$P(X \leq 4) = \sum_{x=0}^{4} b(x;20, 0.1) = 0.9568$$

แปลง X (ตัวแปรสุ่มทวินาม) เป็น X' (ตัวแปรสุ่มปกติ) ดังแสดงในรูปที่ 5.5

$$\begin{aligned} P(X \leq 4) &= P(X' \leq 4.5) \\ &= P\left(Z \leq \frac{4.5 - 20 \times 0.1}{\sqrt{20 \times 0.1 \times 0.9}}\right) \\ &= P(Z \leq 1.86) \\ &= 0.9586 \\ \therefore \text{ ความผิดพลาด} &= 0.9586 - 0.9568 \\ &= 0.0018 \end{aligned}$$

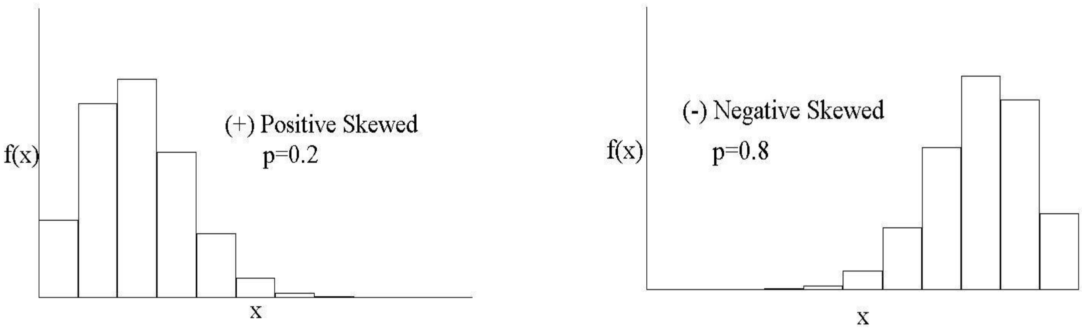


รูปที่ 3.6 การใช้การแจกแจงแบบปกติประมาณค่าการแจกแจงแบบทวินาม

หมายเหตุ

ถ้า $p < \frac{1}{2}$ โค้งจะเบ้ไปทางขวา (Right Skewed หรือ Positive Skewed)

$p > \frac{1}{2}$ โค้งจะเบ้ไปทางซ้าย (Left Skewed หรือ Negative Skewed) ดังรูปที่ 5.6



รูปที่ 3.7 การแจกแจงแบบไม่ปกติ

ตัวอย่างที่ 3.17 ในการจัดนิสิตอยู่หอพักอย่างสุ่มตึกละ 180 คน ถ้า 1/6 ของนิสิตมาจากต่างจังหวัด จงหาความน่าจะเป็นที่นิสิตในหอพักจะมีนิสิตต่างจังหวัดอย่างน้อย 1 ใน 5

วิธีทำ

ให้ X = จำนวนนิสิตต่างจังหวัดในจำนวนนิสิตในหอพัก 180 คน ซึ่งเป็นตัวแปรสุ่มไม่ต่อเนื่องแบบทวินามมี $n = 180, p = \frac{1}{6}$

$$P(X \geq \frac{1}{5} \times 180) = \sum_{x=36}^{180} b(x ; 180, \frac{1}{6})$$

ให้คำนวณโดยใช้ Normal Approximation to Binomial

$\mu \quad = \quad np = 180 \times \frac{1}{6} = 30$

$\sigma \quad = \quad \sqrt{180 \times \frac{1}{6} \times \frac{5}{6}} = \sqrt{25} = 5$

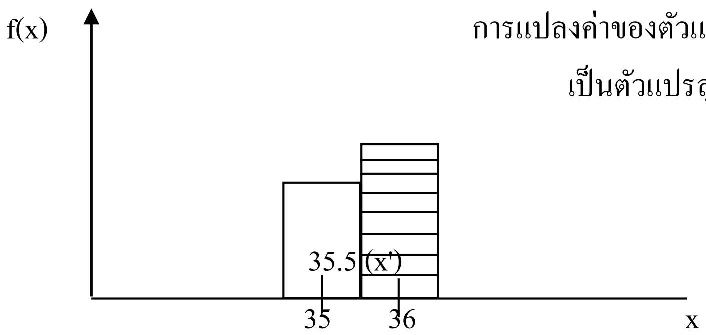
$P(X \geq \frac{1}{5} \times 180) = P(X \geq 36)$

$= P(X' \geq 35.5)$

$= P(Z \geq \frac{35.5-30}{5})$

$= P(Z \geq 1.1)$

$= 0.1357$



การแปลงค่าของตัวแปรสุ่มไม่ต่อเนื่อง X เป็นตัวแปรสุ่มต่อเนื่อง X'

ตัวอย่างที่ 3.18 บริษัทผลิตยาทราบว่าโดยเฉลี่ยมียาที่ใช้ไม่ได้ 5% จงหาความน่าจะเป็นที่ยาใช้ไม่ได้น้อยกว่า 10 เม็ดใน 200 เม็ด

วิธีทำ

ให้ X = จำนวนที่ใช้ไม่ได้ในยา 200 เม็ด เป็นตัวแปรสุ่มแบบทวินาม

$P(X < 10) = \sum_{x=0}^{9} b(x\,;200,0.05)$

ให้คำนวณโดยใช้ Normal Approximation to Binomial

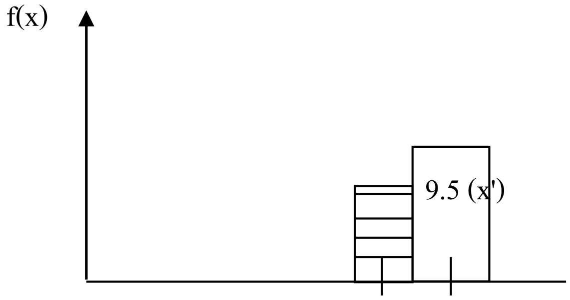
$\mu \quad = \quad 200 \times \frac{1}{20} = 10$

$\sigma \quad = \quad \sqrt{200 \times \frac{1}{20} \times \frac{19}{20}} = 3.08$

$P(X < 10) = P(Z < \frac{9.5-10}{3.08})$

$= P(Z < -0.1623)$

$= 0.4355$



การแปลงค่าของตัวแปรสุ่มไม่ต่อเนื่อง X เป็นตัวแปรสุ่มต่อเนื่อง X'

9 10 x

ตัวอย่างที่ 3.19 ผู้ผลิตยาอ้างว่ายารักษาโรคได้ 80% ถ้าทำการทดลองใช้ยากับผู้ป่วย 100 คน และกำหนดว่าต้องรักษาได้ 75 รายหรือมากกว่านั้น จึงจะยอมรับว่ายามีสรรพคุณตามที่บริษัทกล่าวอ้าง

วิธีทำ

ให้ X = จำนวนผู้ป่วยที่ยารักษาได้จากจำนวนผู้ป่วยที่ทดสอบ 100 คน เป็นตัวแปรสุ่มแบบทวินาม

ใช้ Normal Approximation to Binomial

$$p = 0.8, \quad n = 100$$

$$\mu = 100 \times 0.8 = 80$$

$$\sigma = \sqrt{100 \times 0.8 \times 0.2} = 4$$

$$
\begin{aligned}
P(X \geq 75) &= P\left(Z \geq \frac{74.5 - 80}{4}\right) \\
&= P(Z \geq -1.375) \\
&= 1-P(Z < -1.375) \\
&= 1-0.0846 \\
&= 0.9154
\end{aligned}
$$

3.3.3 การแจกแจงแบบแกมม่า (Gamma Distribution)

การแจกแจงแบบแกมม่ามักจะบิดเบี้ยว (Skewed) ไม่สมมาตรรอบรจุดศูนย์กลาง โดยค่าของตัวแปรสุ่มแบบแกมม่าจะมีค่าเป็นบวกเสมอ ถือเป็นการแจกแจงที่สำคัญและใช้กันมากในทางวิศวกรรมโดยเฉพาะทางด้านอุทกวิทยา

ถ้า X = ตัวแปรสุ่มแบบแกมม่า

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad \text{เมื่อ} \quad 0 \leq x < \infty \quad \text{................(3.32)}$$

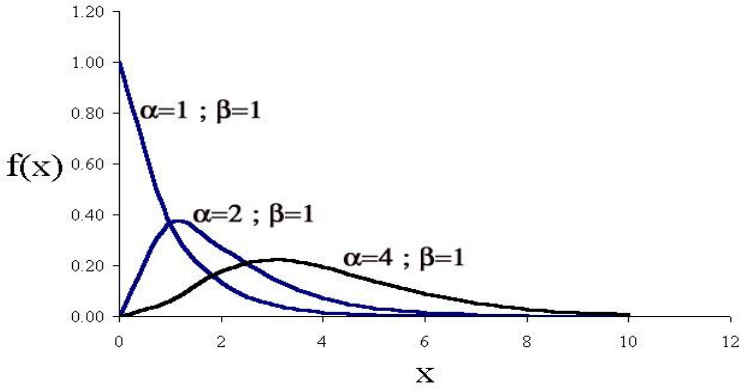
$$\alpha > 0$$

$$\beta > 0$$

เมื่อ α = พารามิเตอร์แสดงรูปร่าง (Shape Parameter)

ถ้า α เปลี่ยนจะมีผลทำให้รูปร่างของการแจกแจงเปลี่ยนดังรูปที่ 3.8

β = พารามิเตอร์แสดงมาตราส่วน (Scale Parameter) ทำหน้าที่เหมือน σ ในการแจกแจงแบบปกติ



รูปที่ 3.8 กราฟการแจกแจงแบบแกมม่าเมื่อ α และ β มีค่าต่างกัน

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx \quad \text{(Gamma Function)} \qquad \text{.....(3.33)}$$

$$\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$$

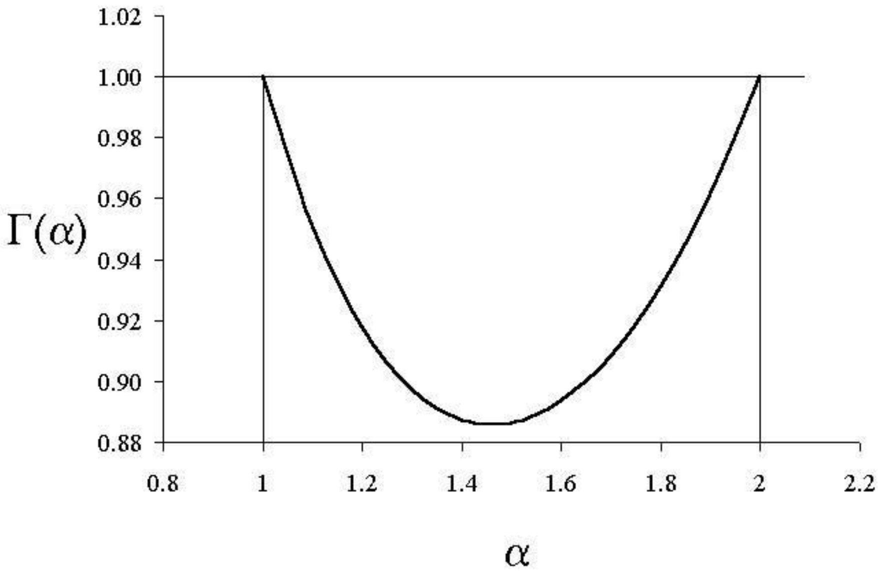
$$= (\alpha-1)\,!$$

$$\Gamma(1) = 1$$

$$\Gamma(2) = 1$$

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

กราฟแสดงว่า Γ (2) แสดงอยู่ในรูปที่ 3.9



รูปที่ 3.9 Gamma Function

ตารางค่า $\Gamma(\alpha)$ เมื่อ $1 \le \alpha < 2$ แสดงอยู่ในตารางที่ 3.2

ตารางที่ 3.2 แกมม่าฟังค์ชั่น (Gamma Function)

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha-1)\Gamma(\alpha-1)$$

n	Γ(n)	n	Γ(n)	n	Γ(n)	n	Γ(n)
1.00	1.00000	1.25	.90640	1.50	.88623	1.75	.91906
1.01	.99433	1.26	.90440	1.51	.88659	1.76	.91237
1.02	.98884	1.27	.90250	1.52	.88704	1.77	.92376
1.03	.98355	1.28	.90072	1.53	.88757	1.78	.92623
1.04	.97844	1.29	.89904	1.54	.88818	1.79	.92877
1.05	.97350	1.30	.89747	1.55	.88887	1.80	.93138
1.06	.96874	1.31	.89600	1.56	.88964	1.81	.93408
1.07	.96415	1.32	.98464	1.57	.89049	1.82	.93685
1.08	.95973	1.33	.89338	1.58	.89142	1.83	.93969
1.09	.95546	1.34	.89222	1.59	.89243	1.84	.94261
1.10	.95135	1.35	.89115	1.60	.89352	1.85	.94561
1.11	.94739	1.36	.89018	1.61	.89468	1.86	.94869
1.12	.94359	1.37	.88931	1.62	.89592	1.87	.95184
1.13	.93993	1.38	.88854	1.63	.89724	1.88	.95507
1.14	.93642	1.39	.88785	1.64	.89864	1.89	.95838
1.15	.93304	1.40	.88726	1.65	.90012	1.90	.96177
1.16	.92980	1.41	.88676	1.66	.90167	1.91	.96523
1.17	.92670	1.42	.88636	1.67	.90330	1.92	.96878
1.18	.92373	1.43	.88604	1.68	.90500	1.93	.97240
1.19	.92088	1.44	.88580	1.69	.90678	1.94	.97610
1.20	.91817	1.45	.88565	1.70	.90864	1.95	.97988
1.21	.91558	1.46	.88560	1.71	.91057	1.96	.98374
1.22	.91311	1.47	.88563	1.72	.91258	1.97	.98768
1.23	.91075	1.48	.88575	1.73	.91466	1.98	.99171
1.24	.90852	1.49	.88595	1.74	.91683	1.99	.99581
						2.00	1.00000

* การแจกแจงแบบเอกซ์โปเนนเชียล (Exponential Distribution) คือการแจกแจงแบบแกมม่า

เมื่อ $\alpha = 1$; $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$ เมื่อ $0 \le x < \infty$ ....(3.34)

* การแจกแจงแบบไคสแควร์ (Chi-Squared Distribution) คือการแจกแจงแบบแกมม่า

เมื่อ $\alpha = \nu/2$; $\beta = 2$ เมื่อ ν(nu) = เลขจำนวนเต็มบวก

$$f(x) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$$ เมื่อ $0 \le x < \infty$ ....(3.35)

การหา μ และ σ^2 ของตัวแปรสุ่มแกมม่า

$$\mu = E(X) = \int_0^\infty x\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}\,dx$$

$$= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty x^\alpha e^{-x/\beta}\,dx$$

ให้ $y = \frac{x}{\beta}$

$dx = \beta dy$

$$\mu = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty \beta^\alpha y^\alpha e^{-y}\beta\,dy$$

$$= \frac{\beta}{\Gamma(\alpha)}\int_0^\infty y^\alpha e^{-y}\,dy$$

$$= \beta\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\beta \qquad \text{.........................(3.36)}$$

$$\sigma^2 = E(X^2)-\mu^2$$

$$E(X^2) = \int_0^\infty x^2\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}\,dx$$

$$= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty x^{\alpha+1}e^{-x/\beta}\,dx$$

ให้ $y = \frac{x}{\beta}$

$dx = \beta dy$

$$E(X^2) = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty \beta^{\alpha+1}y^{\alpha+1}e^{-y}\beta\,dy$$

$$= \frac{\beta^2}{\Gamma(\alpha)}\int_0^\infty y^{\alpha+1}e^{-y}\,dy$$

$$= \beta^2\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)}$$

$$= \beta^2(\alpha+1)\alpha$$

$$\sigma^2 = \beta^2(\alpha+1)\alpha-\beta^2\alpha^2$$

$$= \beta^2\alpha^2+\beta^2\alpha-\beta^2\alpha^2$$

$$= \alpha\beta^2 \qquad \text{.................................(3.37)}$$

ตัวอย่างที่ 3.20 ให้ X = ตัวแปรสุ่มซึ่งมี pdf ดังนี้

$$f(x) = \frac{1}{4}x\,e^{-\frac{x}{2}} \qquad \text{เมื่อ } x \ge 0$$

จงหา μ และ σ^2

วิธีทำ จาก pdf ที่กำหนดให้จะสามารถเขียนในรูปของการแจกแจงแบบแกมม่าได้ดังนี้

$$f(x) = \frac{1}{2^2\,\Gamma(2)}x^{2-1}e^{-\frac{x}{2}} \qquad \text{เมื่อ } x \ge 0$$

$$\alpha = 2$$

$\beta \quad = \quad 2$

$\mu \quad = \quad \alpha\beta = 4$

$\sigma^2 \quad = \quad \alpha\beta^2 = 2x2^2 = 8$

ตัวอย่างที่ 3.21 ปริมาณการใช้น้ำประจำวันของเมือง ๆ หนึ่งมีการแจกแจงแบบแกมม่าโดยมี $\alpha = 2$, $\beta = 3$ ถ้าระบบประปาของเมืองมีถังประปาซึ่งเก็บน้ำได้วันละ 9 ล้านลิตร จงหาความน่าจะเป็น (ความเสี่ยง) ที่น้ำไม่พอใช้

วิธีทำ ให้ X = ปริมาณการใช้น้ำประจำวันของเมือง มีการแจกแจงแบบแกมม่า

$\mu \quad = \quad \alpha\beta = 2x3 = 6$

$\sigma^2 \quad = \quad \alpha\beta^2 = 2x3^2 = 18$

$$P(X > 9) \quad = \quad 1\text{-}P(X \le 9)$$

$$= \quad 1\text{-}\int_0^9 \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha\text{-}1} e^{-\frac{x}{\beta}} dx$$

$$= \quad 1\text{-}\int_0^9 \frac{1}{3^2\,\Gamma(2)} x^{2\text{-}1} e^{-\frac{x}{3}} dx$$

$$= \quad 1\text{-}\frac{1}{9}\int_0^9 -3x \; e^{-\frac{x}{3}} d\left(-\frac{x}{3}\right)$$

$$= \quad 1+\frac{3}{9}\int_0^9 x \, d\left(e^{-\frac{x}{3}}\right)$$

Integrate By-Part

$$= \quad 1+\frac{1}{3}\left[xe^{-\frac{x}{3}}\Big|_0^9 - \int_0^9 e^{-\frac{x}{3}} dx\right]$$

$$= \quad 1+\frac{1}{3}\left[9e^{-3} + \int_0^9 3d\left(e^{-\frac{x}{3}}\right)\right]$$

$$= \quad 1+\frac{1}{3}[9e^{-3} + 3e^{-3} - 3]$$

$$= \quad 1+\frac{1}{3}[12e^{-3} - 3]$$

$$= \quad 4e^{-3} = 0.1991$$

3.3.4 การแจกแจงแบบเอกซ์โปเนนเชียล (Exponential Distribution)

การแจกแจงแบบเอกซ์โปเนนเชียลคือ การแจกแจงแบบแกมม่า กรณีที่ $\alpha = 1$ ดังรูปที่ 3.10 ปกติใช้อธิบายเกี่ยวกับตัวแปรสุ่มดังต่อไปนี้

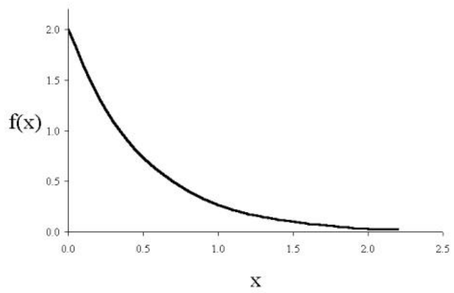
- ระยะเวลาของการรอคอย Arriral time หรือ Wating time หรือ Queueing ของ Service Counter

- อายุการใช้งานของอุปกรณ์ (ระยะเวลาที่อุปกรณ์จะเสีย)
- ระยะเวลาที่ฝนทิ้งช่วง

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \quad \text{เมื่อ} \quad 0 \le x < \infty, \quad \beta > 0 \qquad \text{.......................(3.38)}$$

$\mu = \beta$

$\sigma^2 = \beta^2$



รูปที่ 3.10 การแจกแจงแบบเอกซ์โปเนนเชียล

การแจกแจงแบบเอกซ์โปเนนเชียลมีส่วนคล้ายกับการแจกแจงแบบปัวซองดังตาราง

	Exponential	Poisson
X	Duration of time for one success/failure	No. of success/unit time
f(x)	$\frac{1}{\beta} e^{-x/\beta} \quad 0 \le x < \infty$	$\frac{\mu^x e^{-\mu}}{x!}$ x = 0, 1,2,.....
μ	ระยะเวลา β นาที (ชม.)/1 ครั้ง	$\frac{1}{\beta}$ ครั้ง/นาที (ชม.)
Probability	P(X $\le$ t นาที) = $\int_0^t \frac{1}{\beta} e^{-x/\beta} dx$ $= -[e^{-x/\beta}]^t_0$ $= 1 - e^{-t/\beta}$	P(X $\ge$ 1 ครั้ง/t นาที) (* $\mu = \frac{t}{\beta}$ ครั้ง/t นาที *) $P(X \ge 1) = 1 - P(X = 0)$ $= 1 - \frac{\mu^0 e^{-\mu}}{0!}$ $= 1 - e^{-\mu}$ $= 1 - e^{-t/\beta}$

ตัวอย่างที่ 3.22 ถ้านักตกปลา ตกปลาได้เฉลี่ย 1.8 ตัว/ชั่วโมง จงหาความน่าจะเป็นที่นักตกปลาจะตกปลาได้ 1 ตัวใน 20 นาที

วิธีทำ

ให้ X = ตัวแปรสุ่มแบบเอกซ์โปเนนเชียล แสดงระยะเวลาการตกปลาเป็นชั่วโมง/ปลา 1 ตัว

$\beta = \mu = \frac{1}{1.8}$ ชั่วโมง/ตัว

P (ตกปลา 1 ตัวใน 20 นาที)

= P(X ≤ $\frac{20}{60}$ ชั่วโมง/ตัว)

= P(X ≤ $\frac{1}{3}$ ชั่วโมง/ตัว)

$$= \int_0^{\frac{1}{3}} \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$

$$= \int_0^{\frac{1}{3}} 1.8 e^{-1.8x} dx$$

$$= e^{-1.8x} \Big|_0^{\frac{1}{3}} = -(e^{-1.8x\frac{1}{3}} - 1)$$

$$= 1-e^{-0.6} = 0.45$$

ถ้าให้ Y = ตัวแปรสุ่มแบบปัวซอง แสดงปลาที่ตกได้ใน 20 นาที

μ_y = 1.8 ตัว/ชั่วโมง = $\frac{1.8}{3}$ = 0.6 ตัว/20 นาที

P (ตกปลาได้ 1 ตัวใน 20 นาที)

= P(Y ≥ 1 ตัว/20 นาที)

= 1-P(Y = 0)

= 1-p(0 ; 0.6)

$$= 1-\frac{\mu^y e^{-\mu}}{y!}$$

$$= 1-\frac{(0.6)^0 e^{-0.6}}{0!}$$

$$= 1-e^{-0.6}$$

= 0.45

ตัวอย่างที่ 3.23 อายุการใช้งานของหลอดไฟฟ้ามีการแจกแจงแบบเอกซ์โปเนนเชียลมี β เท่ากับ 1 ถ้าทำการทดสอบหลอดไฟฟ้า 6 หลอด จงหาความน่าจะเป็นที่อย่างมาก 2 หลอดยังใช้การได้หลัง 2.3 ปี

วิธีทำ

t = อายุการใช้งานของหลอดไฟฟ้า

β = 1

$P(t > 2.3 \text{ ปี}) = 1-\int_0^{2.3} (1)e^{-t}\,dt$

$= 1-[-e^{-t}]_0^{2.3}$

$= 1+[e^{-2.3}-1]$

$= e^{-2.3} = 0.1$

โอกาสที่แต่ละหลอดจะมีอายุเกิน 2.3 ปี เท่ากับ 0.1

ให้ X = ตัวแปรสุ่มแบบทวินามแสดงจำนวนหลอดไฟฟ้าที่ใช้ได้หลัง 2.3 ปี

$p = 0.1$

$n = 6$

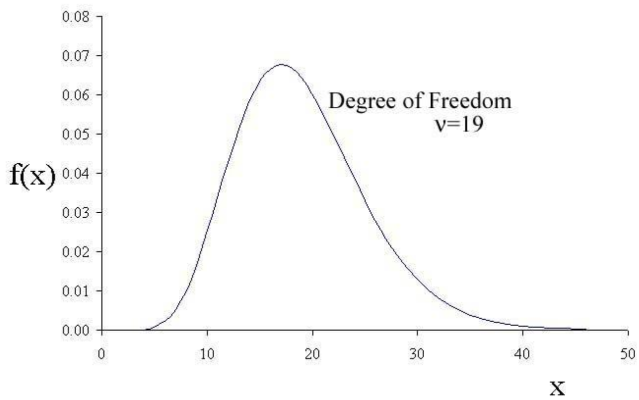
$P(X \le 2) = \sum_{x=0}^{2} b(x\,;6, 0.1)$

$= 0.9841$

3.3.5 การแจกแจงแบบไคสแควร์ (Chi-Squared (χ^2) Distribution)

การแจกแจงแบบแกมม่าซึ่งมี $\alpha = \frac{\nu}{2}$ และ $\beta = 2$ เรียกว่าการแจกแจงแบบไคสแควร์ ดังรูปที่ 3.11

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2} = \frac{\text{Sample Variance}}{\text{Population Variance}} \qquad \text{...(3.39)}$$



รูปที่ 3.11 การแจกแจงแบบไคสแควส์

นิยมใช้ในการทดสอบสมมติฐานเกี่ยวกับความแปรปรวนของประชากรชุดเดียวและการทดสอบสมมติฐานเกี่ยวกับการทดลองหรือการสุ่มตัวอย่างต่าง ๆ ดังนี้

- การทดสอบความเหมาะสมของการแจกแจง (Goodness of Fit)
- การทดสอบความเป็นอิสระ (Independence)
- การทดสอบสัดส่วน (Proportion)
- การทดสอบความเหมือนกันของประชากร (Homogenuity)

รายละเอียดการทดสอบต่าง ๆ จะได้กล่าวถึงในบทที่ 9

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu}{2}-1} e^{-x/2} \quad \text{เมื่อ } 0 \le x < \infty \qquad ....(3.40)$$

เมื่อ ν (nu) = องศาเสรี (Degree of Freedom) = n-1 (เมื่อ n คือจำนวนข้อมูล)

ถ้า ν น้อย การแจกแจงจะเป็นแบบเบ้ขวา (+)

ถ้า ν มาก การแจกแจงจะเข้าใกล้การแจกแจงแบบปกติ (Normal Distribution)

$$\mu = \alpha\beta = \frac{\nu}{2} 2 = \nu$$

$$\sigma^2 = \alpha\beta^2 = \frac{\nu}{2}(2)^2 = 2\nu$$

ค่า χ^2 วิกฤตที่ความน่าจะเป็นต่าง ๆ แสดงอยู่ในตารางที่ A4 ท้ายเล่ม

ตัวอย่างที่ 3.24 บริษัทผู้ผลิตแว่นตาสั่งซื้อกระจกเลนซ์ จากบริษัทผู้ผลิตแห่งหนึ่งซึ่งอ้างว่ามีการควบคุมคุณภาพการผลิตที่ดี ดัชนีการหักเหแสงของเลนซ์แต่ละอันใกล้เคียงกัน มี σ^2 เท่ากับ 1.26×10^{-4} ถ้าจากผลการสุ่มตัวอย่างเลนซ์ 20 ชิ้น พบว่าค่าความเบี่ยงเบนมาตราฐาน S^2 เท่ากับ 2.0×10^{-4} จงคำนวณหาค่า χ^2 และความน่าจะเป็นที่ χ^2 จะมากกว่าค่าที่คำนวณได้

วิธีทำ $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} = \dfrac{(20-1)x2x10^{-4}}{1.26x10^{-4}}$

$= 30.2$

จากตาราง ที่ A4

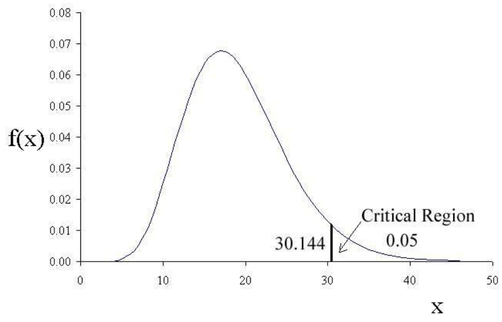
$P(\chi^2 \ge 30.144) = 0.05$ เมื่อ $\upsilon = 20-1 = 19$

$\therefore P(\chi^2 \ge 30.2) < 0.05$

แสดงว่าถ้าคำกล่าวอ้างของบริษัทผู้ผลิตเลนซ์เป็นจริง ($\sigma^2 = 1.26 \times 10^{-4}$) แล้ว มีความน่าจะเป็นเพียงไม่ถึงกว่า 5% ที่จะสุ่มตัวอย่างได้ $S^2 \ge 2.0 \times 10^{-4}$ ซึ่งจะเห็นได้ว่าโอกาสน้อยมาก

ในทางกลับกัน ถ้าตั้งสมมติฐานว่า $\sigma^2 = 1.26 \times 10^{-4}$ สุ่มตัวอย่างได้ $S^2 = 2.0 \times 10^{-4}$ สามารถนำมาทดสอบว่า S^2 ต่างจาก σ^2 อย่างมีนัยสำคัญหรือไม่ โดยการกำหนดนัยสำคัญของการ

ทดสอบ (α) เช่น $\alpha = 0.05$ ถ้าผลการคำนวณ χ^2 พบว่าค่าที่คำนวณได้ตกอยู่ในช่วงวิกฤตดังรูปที่ 5.12 จะปฏิเสธสมมติฐานว่าดัชนีหักเหแสงของเลนซ์มี $\sigma^2 = 1.26 \times 10^{-4}$ ซึ่งอาจทำให้บริษัทผู้ผลิตแว่นตาไม่รับหรือส่งคืนกระจกเลนซ์ดังกล่าวเนื่องจากคุณภาพการหักเหแสงของเลนซ์แต่ละอันต่างกันมากเกินกว่าที่บริษัทกำหนดไว้ จากตัวอย่าง χ^2 ที่คำนวณได้ (30.2) > χ^2 วิกฤต (30.144)



รูปที่ 3.12 การกำหนดช่วงวิกฤตในการทดสอบสมมติฐานด้วย χ^2

รายละเอียดการทดสอบสมมติฐานจะกล่าวถึงในบทที่ 5

3.3.6 การแจกแจงแบบไวบูล (Weibull Distribution)

การแจกแจงแบบไวบูล นิยมใช้อธิบายการแจกแจงความน่าจะเป็นของอายุการใช้งานของอุปกรณ์หรือระบบ คล้ายการแจกแจงแบบเอกซ์โปเนนเชียล และนิยมใช้ในการวิเคราะห์ความเชื่อถือได้ (Reliability) หรือความเสี่ยง (Risk) ในการใช้อุปกรณ์นั้น ๆ

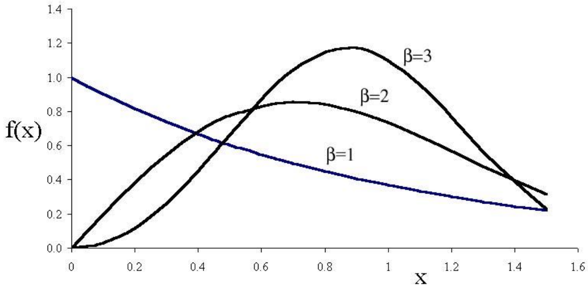
ให้ X = ตัวแปรสุ่มแบบไวบูล

$$f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} \quad \text{เมื่อ} \quad 0 \le x < \alpha \quad \text{.................(3.41)}$$

$$\alpha > 0$$

$$\beta > 0$$

β = พารามิเตอร์แสดงรูปร่างของการแจกแจง ถ้า β เปลี่ยนจะทำให้รูปร่างของการแจกแจงเปลี่ยนดังรูปที่ 5.13 เหมือนพารามิเตอร์ α ของการแจกแจงแบบแกมม่า



รูปที่ 3.12 การแจกแจงแบบไวบูลเมื่อ β ต่าง ๆ กัน และ $\alpha = 1$

ถ้า $\beta = 1$; $f(x) = \alpha e^{-\alpha x}$ = (Exponential)

ถ้า $\beta > 1$; Approximate Normal

$$\mu = \alpha^{-\frac{1}{\beta}} \Gamma\left(1+\frac{1}{\beta}\right) \qquad \text{.....(3.42)}$$

$$\sigma^2 = \alpha^{-\frac{2}{\beta}} \{\Gamma(1+\frac{2}{\beta})-(\Gamma(1+\frac{1}{\beta}))\}^2 \qquad \text{.....(3.43)}$$

➢ **ความน่าเชื่อถือได้ (Reliability)**

ให้ X = เป็นตัวแปรสุ่มแสดงอายุการใช้งานของอุปกรณ์

ถ้าอุปกรณ์ชิ้นหนึ่งมีอายุการใช้งานเท่ากับ t ปี

- อุปกรณ์ชิ้นนั้นจะมีความเชื่อถือได้สูง (Reliable) ถ้า $P(X > t)$มีค่ามาก
- และอุปกรณ์ชิ้นนั้นจะมีความเชื่อถือได้ต่ำ (Unreliable หรือ Risk) ถ้า $P(X < t)$ มีค่ามาก

Reliability ของอุปกรณ์ที่มีอายุการใช้งาน t ปี

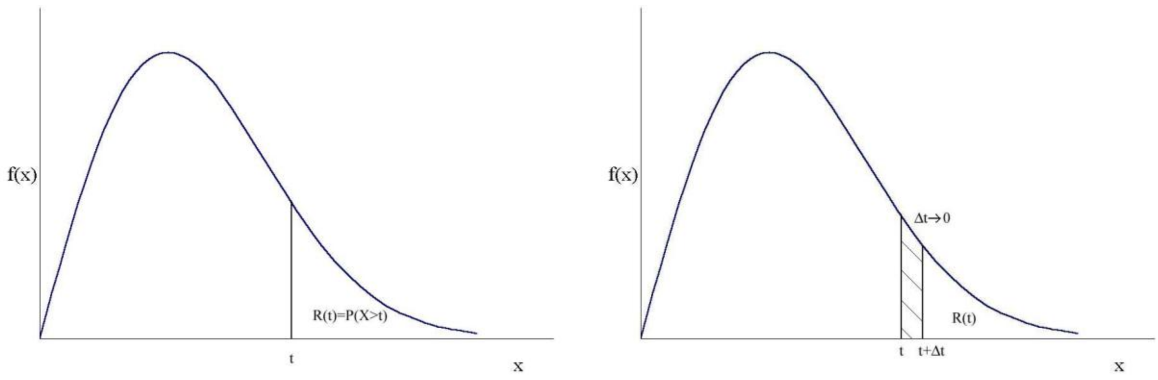
$= R(t) = P(X>t)$ (ดูรูปที่ 5.14)

$= 1-P(X\leq t)$

$= 1-\int_0^t f(x)\,dx$

$= 1-F(t)$ .....(3.44)

ตามทฤษฎีของความน่าเชื่อถือ (Reliablity) ความน่าจะเป็นที่อุปกรณ์จะเสีย (Fail) ในช่วงเวลา $t \longrightarrow t+\Delta t$ ใด ๆ หมายความว่าชิ้นส่วนนั้นใช้งานมาเป็นเวลา t แล้ว เป็นการแจกแจงแบบมีเงื่อนไข (Conditional Probability) หรือพูดอีกนัยหนึ่ง P(X>t) หมายความว่า ชิ้นส่วนจะเสีย (Fail) หลังเวลา X=t หรือ เสียในช่วง $t \longrightarrow t+\Delta t$



รูปที่ 3.14 การกำหนดค่า Reliability

ให้

$$z(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} \times \frac{1}{R(t)}$$

$$= \frac{1}{R(t)} \cdot \frac{d\,F(t)}{dt} = \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)}$$

$$= \text{Failure Rate}$$

$$R(t) = 1-F(t)$$

$$\frac{dR(t)}{dt} = -\frac{dF(t)}{dt} = -f(t)$$

$$\therefore z(t) = -\frac{1}{R(t)} \frac{dR(t)}{dt}$$

$$= -\frac{d}{dt} \ln [R(t)]$$

$$d \ln [R(t)] = -z(t)\, dt$$

$$\ln [R(t)] \Big|_0^t = -\int_0^t z(t)\, dt$$

$$\ln [R(t)] - \ln [R(0)] = -\int_0^t z(t)\, dt \quad ; \{\ln [R(0)] = \ln 1 = 0\}$$

$$\ln [R(t)] = -\int_0^t z(t)\, dt$$

$$R(t) = e^{-\int_0^t z(t)\, dt}$$

3.3.7 สรุปคุณสมบัติที่สำคัญของการแจกแจงตัวแปรสุ่มต่อเนื่อง

คุณสมบัติที่สำคัญของตัวแปรสุ่มต่อเนื่องแบบต่าง ๆ แสดงอยู่ในตารางที่ 3.3

ตารางที่ 3.3 สรุปคุณสมบัติที่สำคัญของการแจกแจงตัวแปรสุ่มต่อเนื่องแบบต่าง ๆ

แบบที่	การแจกแจง	ค่าของ X	f(x)	μ	σ^2	หมายเหตุ
1	แบบเอกรูป หรือ แบบสม่ำเสมอ	$a \le x < b$	$\frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	
2	แบบปกติ	$-\infty \le x < \infty$	$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	μ	σ^2	
3	แบบปกติ มาตราฐาน	$-\infty \le z < \infty$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$	$\mu_z=0$	$\sigma_z^2=1$	
4	แบบแกมม่า	$0 \le x < \infty$	$\frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$ (เมื่อ $\alpha > 0, \beta > 0$)	$\alpha\beta$	$\alpha\beta^2$	
5	แบบเอกซ์โปเนนเชียล	$0 \le x < \infty$	$\frac{1}{\beta} e^{-\frac{x}{\beta}}$ (เมื่อ $\beta > 0$)	β	β^2	คือแกมม่า เมื่อ $\alpha=1$
6	แบบไคสแควร์ (χ^2)	$0 \le x < \infty$	$\frac{1}{2^{\upsilon/2}\Gamma(\upsilon/2)} x^{\frac{\upsilon}{2}-1} e^{-\frac{x}{2}}$ (เมื่อ $\upsilon = n-1$)	υ	2υ	คือแกมม่า เมื่อ $\alpha=\frac{\upsilon}{2}$ $\beta = 2$
7	แบบไวบูล	$x > 0$	$\alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}$ (เมื่อ $\alpha > 0, \beta > 0$)	$\alpha^{-\frac{1}{\beta}}\Gamma(1+\frac{1}{\beta})$	$\alpha^{-\frac{2}{\beta}}\{\Gamma(1+\frac{2}{\beta})-\Gamma(1+\frac{1}{\beta})\}^2$	ถ้า $\beta=1$; ไวบูล=เอกซ์โปเนนเชียล ถ้า $\beta > 1$; ไวบูล $\approx$ ปกติ

3.4 แบบฝึกหัด

1. ระบบสื่อสารในเรือเดินสมุทรลำหนึ่ง ประกอบด้วยชิ้นส่วนที่สำคัญ 4 ชิ้น วิศวกรผู้ออกแบบได้ออกแบบให้ระบบสื่อสารดังกล่าวทำงานได้ตามปกติ ถ้ามีชิ้นส่วนอย่างน้อย 2 ชิ้นใน 4 ชิ้นทำงาน ชิ้นส่วนแต่ละชิ้นทำงานเป็นอิสระต่อกันโดยมีความน่าจะเป็นที่ชิ้นส่วนจะทำงานได้ตามปกติเท่ากับ 0.6 จงหาความน่าจะเป็นที่ระบบสื่อสารในเรือเดินสมุทรดังกล่าว จะทำงานได้ตามปกติ

2. สายการบินแห่งหนึ่งพบว่า 50% ของลูกค้าที่สำรองที่นั่งล่วงหน้าจะมาแสดงสิทธิ์ในการเดินทาง สายการบินมีนโยบายให้สำรองที่นั่งได้ 20 ที่นั่ง สำหรับเที่ยวบินที่รับผู้โดยสารได้ 15 ที่นั่ง จงหาความน่าจะเป็นที่ผู้มาแสดงสิทธิ์จะได้เดินทาง

3. เครื่องยนต์ของเครื่องบินแต่ละลำทำงานเป็นอิสระต่อกัน เครื่องบินจะบินได้อย่างปลอดภัย ถ้าอย่างน้อยครึ่งหนึ่งของเครื่องยนต์ทำงานตามปกติ ถ้าความน่าจะเป็นที่เครื่องยนต์ทำงานปกติเท่ากับ p จงหาว่า p ต้องมีค่าเท่าไร จึงจะทำให้เครื่องบิน 4 เครื่องยนต์มีความปลอดภัยมากกว่าเครื่องบิน 2 เครื่องยนต์

4. ในการศึกษาการถ่ายทอดทางพันธุกรรมของมนุษย์ซึ่งขึ้นอยู่กับยีนคู่หนึ่งกำหนดให้ d คือยีนที่มีลักษณะเด่น และ r คือยีนที่มีลักษณะด้อย

   ผู้ที่มีลักษณะเด่นสมบูรณ์ คือ ผู้ที่มียีนคู่นี้เป็น dd
   ผู้ที่มีลักษณะด้อยสมบูรณ์ คือ ผู้ที่มียีนคู่นี้เป็น rr
   ผู้ที่มีลักษณะเป็นลูกผสม คือ ผู้ที่มียีนคู่นี้เป็น rd

   ถ้าสามีภรรยาคู่หนึ่งซึ่งมีลักษณะยีนคู่นี้เป็น rd ทั้งคู่มีบุตร 4 คน จงหาความน่าจะเป็นที่สามีภรรยาคู่นี้จะมีลูก 3 คน ที่ได้รับยีนที่มีลักษณะเด่น

5. จากสถิติจราจรของถนนสายหนึ่ง พบว่ามีอัตราการเกิดอุบัติเหตุเฉลี่ย 3 ครั้งต่อวัน จงหาความน่าจะเป็นที่จะไม่มีอุบัติเหตุเกิดขึ้นในวันนี้

6. สถิติคนงานที่ประสบอุบัติเหตุจากเครื่องจักร ในช่วง 9 ปี ของโรงงานแห่งหนึ่งพบว่าความน่าจะเป็นที่เกิดอุบัติเหตุในช่วง 3 ปีใด ๆ มีค่าเท่าไร ถ้าคนงานคนหนึ่งประสบอุบัติเหตุ 6 ครั้งในช่วง 9 ปี จงหาความน่าจะเป็นที่จะไม่เกิดอุบัติเหตุเกินกว่า 2 ครั้ง ในช่วง 3 ปีใด ๆ

7. ในการผลิตอาหารกระป๋องชนิดหนึ่ง สมมติว่าอาหารแต่ละกระป๋องจะต้องผ่านกระบวนการตรวจสอบโดบใช้เครื่องจักร 3 ชนิด คือ A, B และ C ตามลำดับทีละกระป๋อง
   ถ้า ความน่าจะเป็นที่อาหารกระป๋องใด ๆ จะไม่ผ่านการตรวจสอบจากเครื่องจักร A คือ 0.3
   ความน่าจะเป็นที่อาหารกระป๋องใด ๆ จะไม่ผ่านการตรวจสอบจากเครื่องจักร B เป็น 0.5
   และ ความน่าจะเป็นที่อาหารกระป๋องใด ๆ จะไม่ผ่านการตรวจสอบจากเครื่องจักร C เป็น 0.2
   อาหารกระป๋องที่ตรงตามมาตรฐานต้องผ่านการตรวจสอบจากเครื่องจักรทั้งสาม (คือต้องผ่านการตรวจสอบจากเครื่องจักร A, B และ C) จงหาความน่าจะเป็นที่จะได้อาหารกระป๋องสำเร็จรูปตรงตามมาตรฐาน 2 กระป๋อง นับตั้งแต่เริ่มเดินเครื่องจักร และผลิตกระป๋องที่ 3 ไม่ตรงตามมาตรฐาน และจงหาความน่าจะเป็นที่จะได้อาหารกระป๋องสำเร็จรูปตรงตามมาตรฐานเป็นกระป๋องที่ 5 ในการตรวจสอบครั้งที่ 8

8. ชาย 3 คน ผลัดกันโยนเหรียญอันหนึ่งคนละครั้ง ถ้าผู้ใดได้ผลลัพธ์จากการโยนเหรียญแตกต่างไปจากอีก 2 คน ผู้นั้นจะต้องเป็นผู้จ่ายค่าอาหาร ถ้าเหรียญขึ้นหน้าเดียวกันทั้ง 3 คน ก็ต้องโยนใหม่อีก จงหาความน่าจะเป็นที่มีการโยนเหรียญทั้งหมดไม่เกิน 3 รอบ

9. จำนวนรถยนต์ที่ผ่านด่านเก็บเงินแห่งหนึ่งระหว่างเวลา 15.00-16.00 น. เฉลี่ยแล้วเป็นจำนวน 300 คัน จงหาความน่าจะเป็นที่จะมีรถยนต์มากกว่า 5 คัน ผ่านด่านเก็บเงินในช่วงเวลาระหว่าง 15.49-15.50 น.

10. เมือง ๆ หนึ่งมีประชากร 10,000 คน ประมาณว่ามีคน 4,000 คน ไม่เห็นด้วยกับการเก็บภาษีแบบใหม่ ถ้าเลือกสุ่มผู้ออกเสียงจำนวน 20 คน และถามความคิดเห็น จงหาความน่าจะเป็นที่อย่างมาก 11 คน เห็นด้วยกับการเก็บภาษีแบบใหม่นี้

11. ถ้า X เป็นตัวแปรสุ่มต่อเนื่องแบบเอกรูป ซึ่งมี Density Function คือ

$$f(x) = \frac{1}{b - a} \quad \text{เมื่อ} \quad a \le x < b$$

   ถ้า $a = 2$ และ $b = 7$ จงคำนวณหา
   (a) $P(X \ge 4)$
   (b) $P(3 \le X < 5.5)$

12. ถ้า X มีการแจงแจกแบบปกติ โดยมีค่าเฉลี่ยเท่ากับ 20 และส่วนเบี่ยงเบนมาตราฐานเท่ากับ 4 จงหา
   ก. $P(X < 10)$

ข. $P(X \geq 25)$

ค. $P(10 \leq X < 25)$

ง. จงหาค่า k ถ้า $P(X < k) = 0.20$

จ. จงหาค่า k ถ้า $P(X > k) = 0.20$

13. ในการสอบวิชาคณิตศาสตร์ พบว่านิสิตได้คะแนนเฉลี่ย 82 คะแนน ส่วนเบี่ยงเบนมาตราฐานเท่ากับ 5 คะแนน นิสิตที่ได้คะแนนระหว่าง 88 ถึง 94 ได้เกรด B ถ้าคะแนนผลการสอบมีการแจกแจงแบบปกติ และมีนิสิต 8 คนได้เกรด B ถามว่ามีนิสิตเข้าสอบวิชานี้กี่คน

14. กำลังในการรับแรงดึงของโลหะชนิดหนึ่งซึ่งมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 10,000 ksc มีค่าส่วนเบี่ยงเบนมาตราฐานเท่ากับ 100 ksc ถ้าการวัดกำลังรับแรงดึงของโลหะวัดได้ละเอียดถึง 50 ksc

ก. จงหา % ของชิ้นโลหะดังกล่าวซึ่งสามารถรับแรงดึงได้มากกว่า 10,150 ksc

ข. ถ้ามาตราฐานการใช้งานของโลหะดังกล่าวกำหนดว่าโลหะต้องรับแรงถึงได้ระหว่าง 9,800 ถึง 10,200 ksc จงคำนวณหาว่ามีชิ้นโลหะดังกล่าวกี่ % ที่ไม่เป็นไปตามมาตราฐานที่กำหนด

15. บริษัทยาพบว่า 5% ของยาคุมกำเนิดชนิดหนึ่งมีตัวยาซึ่งต่ำกว่ามาตราฐาน และมีผลทำให้การคุมกำเนิดไม่ได้ผล จงคำนวณหาความน่าจะเป็นที่มียาคุมกำเนิดน้อยกว่า 10 ชุด ไม่ได้ผลจากผลการทดสอบยา 200 ชุด

16. จากสถิติตำรวจจราจรในเมือง ๆ หนึ่งพบว่า โดยเฉลี่ยผู้ขับขี่รถยนต์ในคืนวันสุดสัปดาห์ 1 คนใน 10 คน ดื่มของมึนเมาขณะขับรถ ถ้าทำการสุ่มตรวจผู้ขับรถในคืนวันเสาร์ 400 คน จงหาความน่าจะเป็นที่จำนวนผู้ดื่มของมึนเมาขณะขับรถ

ก. น้อยกว่า 32 คน

ข. มากกว่า 49 คน

ค. อย่างน้อย 35 คนแต่น้อยกว่า 47 คน

17. ในเมือง ๆ หนึ่ง ปริมาณการใช้ไฟฟ้าประจำวันซึ่งมีหน่วยเป็นล้านกิโลวัตต์-ชั่วโมง เป็นตัวแปรสุ่มแบบแกมม่ามีค่าเฉลี่ย $\mu = 6$ และค่าความแปรปรวน $\sigma^2 = 12$

ก. จงหา α, β

ข ความน่าจะเป็นที่ปริมาณการใช้จะมากกว่า 12 ล้านกิโลวัตต์-ชั่วโมง ต่อวัน

18. อายุการใช้งานของอุปกรณ์ไฟฟ้าชนิดหนึ่งมีการแจกแจงแบบเอกซ์โปเนนเชียล (Exponential) โดยมีอายุเฉลี่ย $\mu = 2$ ปี ถ้ามีการติดตั้งอุปกรณ์ไฟฟ้าดังกล่าวในโรงงานแห่งหนึ่งจำนวน 150 ชุด จงหาความน่าจะเป็นซึ่งอุปกรณ์ไฟฟ้าดังกล่าวเสียในช่วงปีแรก อย่างน้อย 45 ชุด

19. ปริมาณการใช้น้ำประจำวัน (ล้านแกลลอน) ของเมืองๆ หนึ่ง มีการแจกแจง แบบแกมมา $\alpha = 3$ และ $\beta = 1$ ถ้าที่เก็บน้ำสำหรับใช้ประจำวันบรรจุได้ 4 ล้านแกลลอน จงหาความน่าจะเป็นที่
    ก. ในวันหนึ่งจะมีน้ำไม่พอใช้
    ข. ปริมาณน้ำที่ใช้มีค่าอยู่ระหว่าง 2.5 ถึง 3.5 ล้านแกลลอนต่อวัน

20. บริษัทผลิตเครื่องคิดเลขยี่ห้อหนึ่งให้ใบรับรอง (Warranty) คุณภาพสินค้าเป็นเวลา 1 ปี ถ้าภายในหนึ่งปีเครื่องคิดเลขเสียไม่ว่าด้วยสาเหตุใดก็ตาม บริษัทจะทำการทดแทนให้ด้วยเครื่องใหม่ ช่วงเวลาอายุการใช้งานของเครื่องคิดเลขรุ่นนี้มีฟังก์ชั่นการแจกแจงแบบเอกซ์โปเนนเซียล

$$f(x) = 0.125\, e^{-0.125X} \quad ; x > 0$$

    ก. จงหาความน่าจะเป็นที่เครื่องคิดเลขนี้จะเสียก่อนที่ใบรับรองจะหมดอายุ
    ข. ถ้าค่าใช้จ่ายในการผลิตเครื่องคิดเลขนี้คิดเป็นเงิน 2,000 บาทต่อเครื่อง และกำไรในการขายแต่ละครั้งเท่ากับ 1,000 บาทต่อเครื่อง กำไรที่ได้จะเปลี่ยนเป็นเท่าใดถ้าใช้นโยบายเปลี่ยนเครื่องใหม่ตามใบรับรองสินค้า

21. ก. ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบเอกซ์โปเนนเชียล และ $P[X > 1] = P[X \leq 1]$ จงหาค่าเฉลี่ย และค่าความแปรปรวนของ X
    ข. ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบไคสแควร์ มีค่าความแปรปรวนเท่ากับ 20 จงหาค่าองศาเสรี
    ค. ถ้าตัวแปรสุ่ม X มีค่าความแปรปรวนเท่ากับ 1.8 และค่าเฉลี่ยเท่ากับ 0.36 X มีการแจกแจงแบบแกมมา จงเขียนฟังก์ชั่นการแจกแจงความน่าจะเป็นของ X
    ง. ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 15 และ $P[X > 9] = 0.9595$ จงหาค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

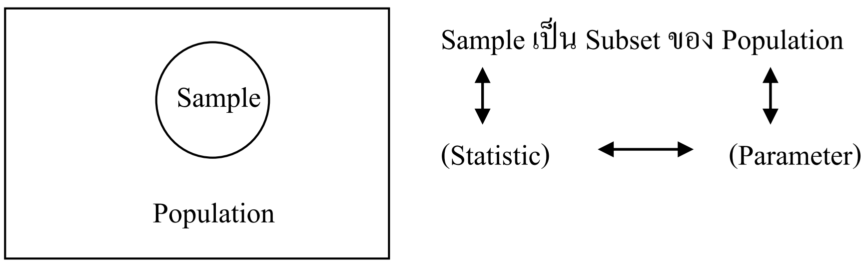
บทที่ 4
การสุ่มตัวอย่างและการประมาณค่า (Random Sampling and Estimation)

4.1 ประชากร (Population)

ประชากร คือ เซทซึ่งประกอบด้วยค่าสังเกตทั้งหมด (Totality of the Observations) ของสิ่งที่กำลังศึกษาหรือทดลอง ซึ่งอาจจะเป็นเรื่องเกี่ยวกับคน สัตว์ พืช หรือสิ่งที่มนุษย์สร้างขึ้นมา จำนวนค่าสังเกตทั้งหมดของประชากรจะบอกขนาดของประชากร เช่น ในการศึกษากลุ่มเลือดของนิสิตในมหาวิทยาลัย ซึ่งมีนิสิต 3,000 คน ขนาดของประชากรคือ 3,000 ประชากรบางชุดเป็นแบบที่นับจำนวนได้ (Finite) เช่น กรณีกลุ่มเลือดนิสิต 3,000 คน หรืออายุคนในอำเภอใดอำเภอหนึ่ง หรือความสูงของคนในจังหวัดใดจังหวัดหนึ่ง แต่ประชากรบางชุดไม่สามารถนับจำนวนได้ (Infinite) เช่นการโยนเหรียญแบบไม่มีที่สิ้นสุดหรือการวัดอุณหภูมิอากาศประจำวันซึ่งจะต้องวัดต่อไปในอนาคตซึ่งถือเป็นประชากรแบบนับจำนวนไม่ได้

ประชากรต่างจากเอกภพสัมพัทธ์ (Sample Space) ตรงที่ประชากรหมายถึง เซทของค่าสังเกตทุกค่า ซึ่งแต่ละค่าจะเหมือนหรือไม่เหมือนกันก็ได้ ขณะที่เอกภพสัมพัทธ์คือ เซทซึ่งแต่ละค่าแตกต่างกัน เช่น กรณีกลุ่มเลือดนิสิต 3,000 คน ประชากรคือกลุ่มเลือดซึ่งอาจเป็น A, A, B, AB, O, O,……..ทั้ง 3,000 ค่า ขณะที่เอกภพสัมพัทธ์ของกลุ่มเลือดนิสิตคือ O, A, B, AB เท่านั้น

ถ้าประชากรมีขนาดใหญ่มากหรือนับไม่ถ้วน การศึกษาประชากรทั้งหมดเป็นเรื่องที่ยากลำบาก เสียเวลา และค่าใช้จ่ายสูง จึงควรใช้วิธีการสุ่มตัวอย่าง(Random Sampling) ขนาดพอสมควรจากประชากร เพื่อศึกษาและหาข้อสรุปเกี่ยวกับประชากรทั้งหมด ***โดยสุ่มจากค่าสังเกตที่เป็นอิสระต่อกัน (Independent) สุ่มแบบไม่เจาะจง(Random)*** แล้วนำมาคำนวณหาค่า ***สถิติ*** เพื่อหาข้อสรุปเกี่ยวกับ ***พารามิเตอร์*** ของประชากรต่อไป



4.2 การสำรวจสำมะโนครัว (Census)

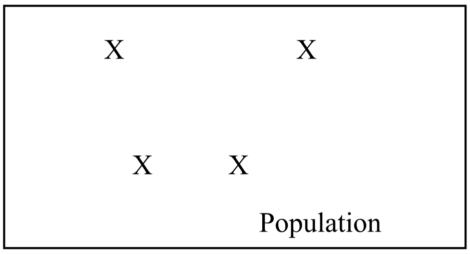
คือ การสำรวจซึ่งเข้าถึงสมาชิกทุกตัวของประชากร (Every Member of Population) ซึ่งเป็นการสำรวจที่เสียเวลาและค่าใช้จ่ายสูงมาก

4.3 วิธีการสุ่มตัวอย่าง (Random Sampling Techniques)

วิธีการสุ่มตัวอย่างมี 4 วิธี คือ

4.3.1 การสุ่มตัวอย่างแบบธรรมดา (Simple Random Sampling)

การสุ่มตัวอย่างแบบธรรมดา คือการสุ่มตัวอย่างซึ่งสมาชิกของประชากรแต่ละตัวมีโอกาสถูกเลือกเท่ากัน ดังรูปที่ 4.1 เช่นการสุ่มโดยวิธีล๊อตเตอรี่ (Lottery) หรือการจับเบอร์ (Drawing) หรือใช้ตารางเลขสุ่ม (Table of Random Number) ดังแสดงในตารางที่ A5 ของภาคผนวก



รูปที่ 4.1 การสุ่มตัวอย่างแบบธรรมดา

ตัวอย่างการสุ่มโดยวิธีล๊อตเตอรี่ ต้องการสุ่มตัวอย่างนักเรียน 5 คน จากชั้นเรียนซึ่งมีนักเรียน 50 คน ทำได้โดยการเขียนชื่อนักเรียนทั้ง 50 คน ลงในแผ่นกระดาษ 50 ชิ้น ใส่ลงในกระป๋อง เขย่าให้ทั่ว แล้วสุ่มหยิบชื่อทีละแผ่น 5 ครั้ง จะได้ตัวอย่างสุ่มแบบธรรมดา ซึ่ง $n = 5$

ตัวอย่างการสุ่มโดยวิธีใช้ตารางเลขสุ่ม ต้องการสุ่มตัวอย่างนิสิต 30 คน จากนิสิตทั้งหมด 400 คน โดยการจัดเรียงชื่อนิสิต 400 คน ตามลำดับตัวอักษร แล้วกำหนดหมายเลขจาก 001-400

หลับตาเอานิ้วจิ้มตัวเลขในตารางเลขสุ่ม (เลขสุ่มเป็นตัวเลข 5 หลัก ซึ่งถูกสร้างจาก Random Process) เช่น ถ้าตัวเลขเป็น **333**62 จะได้นิสิตที่สุ่มคนแรกเบอร์ 333 (เลข 3 หลักแรกของเลขสุ่ม) เลือกตัวเลขสุ่มในแถวเดียวกัน ลำดับถัดไป (**949**04) ปรากฏว่าไม่มีนิสิตหมายเลข 949 ให้เลือก

หมายเลขสุ่มลำดับถัดไป (**312**73) จะได้นิสิตหมายเลข 312 เป็นคนที่ 2 เลือกทำนองเดียวกันจนครบ 30 คน ที่ต้องการ ดูรูปที่ 4.2

51821	51259	77452	16308	60756	92144	49442
52404	60268	89368	19885	55322	44819	01188
33362	**949**04	**312**73	04146	18594	29852	71585
46369	58586	23216	14513	83149	98736	23495
33787	09998	42698	06691	76988	13602	51851

รูปที่ 4.2 การสุ่มโดยใช้ตารางเลขสุ่ม

4.3.2 การสุ่มตัวอย่างโดยการแบ่งเป็นพวก (Stratified Random Sampling)

แบ่งประชากรออกเป็นกลุ่มย่อย (Subpopulation) หรือเป็นพวก (Strata) เช่นแบ่งตามระดับรายได้ เพศ วัย หรือระดับการศึกษา ทำการสุ่มตัวอย่างแบบธรรมดาในแต่ละกลุ่มย่อย (แต่ละพวก) ดังรูปที่ 4.3 จะได้จำนวนตัวอย่างตามจำนวนกลุ่มย่อย (เช่น 3 ตัวอย่าง จาก 3 กลุ่มย่อย) เมื่อนำมารวมกันจะได้ตัวอย่างที่ต้องการซึ่งเรียกว่า Stratified Random Sample ขนาดของตัวอย่างที่เลือกจากแต่ละกลุ่มย่อยจะแปรผันตามขนาดของประชากรในกลุ่มย่อย

A	B	C
X X X	X X X	X X X
รายได้ ต่ำ	รายได้ ปานกลาง	รายได้ สูง

รูปที่ 4.3 การสุ่มตัวอย่างโดยการแบ่งเป็นพวก

4.3.3 การสุ่มตัวอย่างแบบมีระบบ (Systematic Random Sampling)

กรณีที่ประชากรใหญ่มาก อาจไม่สะดวกที่จะทำการสุ่มแบบธรรมดา เช่นต้องการสุ่ม 150 ครัวเรือน จาก 45,000 ครัวเรือน จึงควรใช้วิธีสุ่มแบบมีระบบ โดยการจัดเรียงรายชื่อตามลำดับตัวอักษร แล้วกำหนดหมายเลขให้แต่ละครัวเรือน เนื่องจากอัตราส่วน ประชากร/ตัวอย่าง = 45,000/150 = 300 จึงเลือกสุ่มค่าสังเกตแรก จาก 300 ครัวเรือนแรก โดยวิธีล๊อตเตอรี่ หรือใช้ ตารางเลขสุ่ม สมมติว่า

ได้ค่าสังเกตแรก เบอร์ 210 จะเลือกค่าสังเกตที่ 2, 3, 4,........,150 แบบมีระบบจากครัวเรือนลำดับที่ 210+300, 210+600, 210+900,.......,210+(150-1) x 300 (ดูรูปที่ 4.4 ประกอบ)

1 X	2 X	3 X	● X	● X	● X	● X
● X	● X	● X	● X	● X	● X	● X
● X	● X	● X	● X	● X	● X	● X
● X	● X	● X	● X	● X	(n-1) X	n X

ประชากร N แบ่งเป็น n กลุ่ม ๆ ละ k (=N/n) ถ้าสุ่มกลุ่มที่ 1 ได้หมายเลข i จะได้ตัวอย่างสุ่มจำนวน n ค่าสังเกตดังนี้ {i, i+k, i+2k,......., i+(n-1)k}

รูปที่ 4.4 การสุ่มตัวอย่างแบบมีระบบ

4.3.4 การสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling)

บางครั้งประชากรกลุ่มเป้าหมาย กระจายตัวอยู่ในพื้นที่ต่าง ๆ จนยากต่อการเข้าสุ่มตัวอย่าง กรณีนี้จะต้องแบ่งประชากรออกเป็นกลุ่มตามพื้นที่ (Geographical Groups หรือ Clusters) ***โดยให้ทุก Cluster มีคุณสมบัติเหมือนกัน*** หรือเป็นตัวแทนของประชากร เลือกสุ่ม k Clusters จาก Clusters ทั้งหมดโดยวิธี Simple Random Sampling แล้วเลือกสุ่มสมาชิกจาก k Clusters ที่เลือก ตามจำนวนที่กำหนด

เช่น ต้องการสุ่ม 1,000 ครัวเรือน จากจังหวัดนครปฐม โดยการแบ่งจังหวัดนครปฐม เป็น 40 เขต (ตำบล) เลือกสุ่ม 5 เขต จาก 40 เขต โดยวิธีสุ่มแบบธรรมดาหลังจากนั้นจึงเลือกสุ่ม 1,000 ครัวเรือน จาก 5 เขตที่เลือกไว้

4.4 ค่าสถิติที่สำคัญ

หลังจากสุ่มตัวอย่างได้แล้ว จะสามารถนำมาคำนวณหาค่าสถิติจากตัวอย่าง เพื่อนำไปอนุมานหรือหาข้อสรุปเกี่ยวกับค่าพารามิเตอร์ของประชากร ค่าสถิติที่สำคัญได้แก่ ค่าที่แสดงแนวโน้มเข้าสู่จุดศูนย์กลางและค่าความแปรปรวน

สถิติ (ตัวอย่าง) $\rightarrow$ พารามิเตอร์ (ประชากร)

4.4.1 แนวโน้มเข้าสู่จุดศูนย์กลางของตัวอย่าง (Central Tendency)

ค่าสถิติที่แสดงแนวโน้มเข้าสู่จุดศูนย์กลาง ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยม

ให้ ตัวอย่าง = $\{x_1, x_2, \ldots\ldots\ldots\ldots, x_n\}$ ซึ่ง x_I แต่ละค่าเป็นอิสระต่อกัน

➢ ค่าเฉลี่ยจากตัวอย่าง (Mean)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.1)$$

$\bar{x}$ = เป็นตัวประมาณค่าของ μ (พารามิเตอร์ของประชากร)

➢ ค่ามัธยฐาน (Median)

$$\underline{x} = x_{\frac{n+1}{2}} \quad \text{ถ้า n เป็นเลขคี่ เช่น n = 5} \qquad \ldots\ldots\ldots\ldots(4.2)$$

$$\underline{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n+1}{2}}}{2} \quad \text{ถ้า n เป็นเลขคู่ เช่น n = 6} \qquad \ldots\ldots\ldots\ldots(4.3)$$

หาโดยการจัดเรียงข้อมูลจากน้อยไปมาก เช่น

ถ้า n = 5 ; $\underline{x} = x_{\frac{5+1}{2}} = x_3$

ถ้า n = 6 ; $\underline{x} = \frac{x_3 + x_4}{2}$

➢ ค่าฐานนิยม (Mode)

คือ ค่าสังเกตของตัวอย่างที่มีความถี่มากที่สุด (ค่าตัวแปรสุ่มที่มีความถี่มากที่สุด)

4.4.2 ความแปรปรวนของตัวอย่าง (Variability)

ค่าสถิติที่แสดงความแปรปรวนของตัวอย่างได้แก่ ค่าความแปรปรวน ค่าส่วนเบี่ยงเบนมาตราฐาน และค่าพิสัยของตัวอย่าง

➢ ค่าความแปรปรวนของตัวอย่าง (Sample variance)

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{เป็นตัวประมาณค่าของ } \sigma^2 \qquad \ldots\ldots\ldots\ldots(4.4)$$

➢ ค่าส่วนเบี่ยงเบนมาตราฐานของตัวอย่าง (Sample standard deviation)

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad \text{เป็นตัวประมาณค่าของ } \sigma \qquad \ldots\ldots\ldots\ldots(4.5)$$

➢ พิสัย (Range) = $x_{max} - x_{min}$ ....(4.6)

สูตรการคำนวณหา S^2

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

$$= \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

$$= \frac{\sum x_i^2 - n\left(\frac{\sum x_i}{n}\right)^2}{n-1}$$

$$= \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)} \qquad \text{..........................(4.7)}$$

4.4.3 ตัวประมาณค่าที่ไม่ลำเอียง (Unbiased Estimator)

ค่าสถิติตัวใดจะเป็นตัวประมาณค่าที่ไม่ลำเอียงก็ต่อเมื่อ Expected Value ของค่าสถิติตัวนั้นมีค่าเท่ากับค่าพารามิเตอร์ที่ต้องการประมาณ หรือ

E(Statistics)	=	Parameter

เช่น

$$E(S^2) = E\left[\frac{\sum(x_i - \bar{x})^2}{n-1}\right]$$

$$= \frac{1}{n-1} E\left[\sum \{(x_i-\mu)-(\bar{x} - \mu)\}^2\right]$$

$$= \frac{1}{n-1} E\left[\sum \{(x_i-\mu)^2-2(\bar{x}-\mu)(x_i-\mu)+(\bar{x}-\mu)^2\}\right]$$

$$= \frac{1}{n-1} E\left[\sum (x_i-\mu)^2-2(\bar{x}-\mu)\sum (x_i-\mu)+\sum (\bar{x}-\mu)^2\right]$$

$$= \frac{1}{n-1} E\left[\sum (x_i-\mu)^2-2(\bar{x}-\mu) (\sum x_i-n\mu)+n (\bar{x}-\mu)^2\right]$$

$$= \frac{1}{n-1} E\left[\sum (x_i-\mu)^2-2n(\bar{x}-\mu)^2+n(\bar{x}-\mu)^2\right]$$

$$= \frac{1}{n-1} E\left[\sum (x_i-\mu)^2-n (\bar{x}-\mu)^2\right]$$

$$= \frac{1}{n-1} \left[\sum E (x_i-\mu)^2-n\, E (\bar{x}-\mu)^2\right] \text{...........*}$$

$$= \frac{1}{n-1} \left[n.\sigma^2-n.\frac{\sigma^2}{n}\right]$$

$$= \frac{1}{n-1} \left[n.\, \sigma^2-\sigma^2\right]$$

$$= \frac{1}{n-1}.\sigma^2(n-1) = \sigma^2 \qquad \text{....................................(4.8)}$$

$$* E(\bar{x}-\mu)^2 = E\left(\frac{\sum x_i}{n} - \mu\right)^2$$

$$= \frac{1}{n^2} E\left(\sum x_i - n\mu\right)^2$$

$$= \frac{1}{n^2} E(x_1+x_2+.......+x_n-n\mu)^2$$

$$= \frac{1}{n^2} E[(x_1-\mu)+(x_2-\mu)+.........+(x_n-\mu)]^2$$

ถ้า $x_1, x_2, ..., x_n$ เป็นอิสระต่อกัน (Independent)

$$E(\bar{x}-\mu)^2 = \frac{1}{n^2}[E(x_1-\mu)^2+E(x_2-\mu)^2+........+E(x_n-\mu)^2]$$

$$= \frac{n.\sigma^2}{n^2} = \frac{\sigma^2}{n} \qquad \text{..................................(4.9)}$$

ดังนั้น $S^2_{\bar{x}}$ เป็นตัวประมาณค่าที่ไม่ลำเลียงของ $\frac{\sigma^2}{n}$

4.5 การแจกแจงของตัวอย่างสุ่ม (Sampling Distributions)

ค่าสถิติที่คำนวณจากตัวอย่าง เช่น ค่าเฉลี่ย $\bar{x}$ ค่าความแปรปรวน (S^2) และค่าสัดส่วน (Proportion, $\hat{p}$) จะมีค่าแตกต่างกันไปตามตัวอย่างที่สุ่มได้ ถ้าตัวอย่างที่สุ่มได้เป็นตัวแทนที่ดีของประชากร ค่าสถิติที่คำนวณจากตัวอย่างนั้นจะเป็นตัวประมาณค่าที่ดีของพารามิเตอร์

สุ่ม	→	สถิติ	→	พารามิเตอร์
(Sampling)		(Statistics)		(Parameters)

ถ้า Sample ดี ; Statistics เป็นตัวประมาณค่าที่ดีของ Parameter

ถ้า Sample ไม่ดี ; Statistics เป็นตัวประมาณค่าที่ไม่ดีของ Parameter

ถ้ามีการสุ่มตัวอย่างหลายครั้ง เช่น n ครั้ง จะได้ค่าสถิติ $\bar{x}_i$, S_i^2 และ $\hat{p}_i$ จำนวนอย่างละ n ค่า ซึ่งแตกต่างกันไป ดังนั้นจึงอาจกล่าวได้ว่าสถิติคือตัวแปรสุ่ม (X) ตัวหนึ่ง ที่มีการแจกแจงความน่าจะเป็นอยู่รอบ ๆ ค่าเฉลี่ยของค่าสถิติตัวนั้น การแจกแจงความน่าจะเป็นของค่าสถิติ เรียกว่า ***การแจกแจงของตัวอย่างสุ่ม (Sampling Distribution)*** มี f(x) ซึ่งแตกต่างกันไปตามขนาดของประชากร ขนาดตัวอย่าง และวิธีการสุ่มตัวอย่าง

การแจกแจงของตัวอย่างสุ่มที่จะกล่าวถึงในบทนี้มี 7 แบบ ได้แก่

(1) การแจกแจงของค่าเฉลี่ยของตัวอย่าง ($\bar{X}$)
(Sampling Distribution of Mean)

(2) การแจกแจงของผลต่างของค่าเฉลี่ยของ 2 ตัวอย่าง ($\bar{X}_1$-$\bar{X}_2$)
(Sampling Distribution of Difference of Means)

(3) การแจกแจงของค่าเฉลี่ยของผลต่างของค่าที่สุ่มเป็นคู่ ($\bar{D}$)
(Sampling Distribution of Mean of Difference of Paired Observation)

(4) การแจกแจงของสัดส่วนของตัวอย่าง $\hat{P}$
(Sampling Distribution of Proportions)

(5) การแจกแจงของผลต่างของสัดส่วนของ 2 ตัวอย่าง ($\hat{P}_1$-$\hat{P}_2$)
(Sampling Distribution of Difference of Proportions)

(6) การแจกแจงของความแปรปรวนของตัวอย่าง (S^2) หรือ χ^2
(Sampling Distribution of Variance or χ^2 Distribution)

(7) การแจกแจงของสัดส่วนของความแปรปรวนของ 2 ตัวอย่างหรือ F
(Sampling Distribution of Proportion of Variances or F Distribution)

4.5.1 การแจกแจงของค่าเฉลี่ยของตัวอย่าง ($\bar{X}$)

(1) กรณีทราบการแจกแจงของประชากร (ทราบ σ^2)

$X \sim n\,(x;\ \mu, \sigma)$

สุ่มตัวอย่างได้ $\rightarrow$ $(x_1, x_2, \dots\dots, x_n) \rightarrow \bar{x} \rightarrow \bar{X}$ เป็นตัวแปรสุ่ม

$$\bar{X} \sim n\,(\bar{x}\ ;\ \mu, \frac{\sigma}{\sqrt{n}})$$

และ $$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad \text{......................................(4.10)}$$

(2) กรณีไม่ทราบการแจกแจงของ X (ไม่ทราบ σ^2)

♦ **ถ้า $n > 30$**

ตามทฤษฎี “Central limit theorem” ซึ่งกล่าวว่าตัวแปรสุ่มที่เกิดจากค่าของตัวแปรสุ่มตัวอื่นหลาย ๆ ค่ามารวมกัน จะมีการแจกแจงแบบปกติ

ไม่ทราบ σ^2 ใช้ S^2 แทน

ดังนั้น $\bar{X} \sim n\,(\bar{x}\ ;\ \mu, \frac{S}{\sqrt{n}})$

และ $z = \dfrac{\bar{x} - \mu}{S / \sqrt{n}}$ ....(4.11)

ตัวอย่างที่ 4.1 สุ่มตัวอย่างจำนวน 16 ตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ ซึ่งมี $\mu = 50$ และ $\sigma = 5$ จงหา $P(\mu_{\bar{x}}-1.9\,\sigma_{\bar{x}} \leq \bar{X} < \mu_{\bar{x}}-0.4\,\sigma_{\bar{x}})$

วิธีทำ

$n = 16;\ \mu = 50;\ \sigma = 5$

$P(\mu_{\bar{x}}-1.9\,\sigma_{\bar{x}} \leq \bar{X} < \mu_{\bar{x}}-0.4\sigma_{\bar{x}})$

$P(\mu-1.9\,\dfrac{\sigma}{\sqrt{n}} \leq \bar{X} < \mu-0.4\dfrac{\sigma}{\sqrt{n}})$

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{5}{\sqrt{16}} = \dfrac{5}{4}$

$P(50-1.9 \times \dfrac{5}{4} \leq \bar{X} < 50-0.4\times\dfrac{5}{4})$

$= P(47.63 \leq \bar{X} < 49.5)$

$= P\left(\dfrac{47.63-50}{\frac{5}{4}} \leq Z < \dfrac{49.5-50}{\frac{5}{4}}\right)$

$= P(-1.9 \leq Z < -0.4)$

$= 0.3446 - 0.0287$

$= 0.3159$

♦ **ถ้า $n < 30$**

$\bar{X}$ จะมีการแจกแจงแบบ t (t Distribution)

$\bar{X} \sim t(\bar{x};\mu, \dfrac{S}{\sqrt{n}})$

$t = \dfrac{\bar{x} - \mu}{S / \sqrt{n}}$ ....……............(4.12)

t มี Degree of Freedom เท่ากับ n-1 รายละเอียดเกี่ยวกับการแจกแจงแบบ t จะแสดงในหัวข้อถัดไป

4.5.2 การแจกแจงแบบ t (t-Distributions)

ให้ T = ตัวแปรสุ่มแบบ t ซึ่งมีการแจกแจงแบบสมมาตรรอบจุดศูนย์กลาง มีค่าเฉลี่ยเท่ากับศูนย์ จึงมีลักษณะคล้ายการแจกแจงแบบปกติมาตราฐาน (Z) แต่ลักษณะของการแจกแจงจะแตกต่างกันไปตาม

ค่าองศาเสรี (Degree of Freedom) ดังรูปที่ 4.5

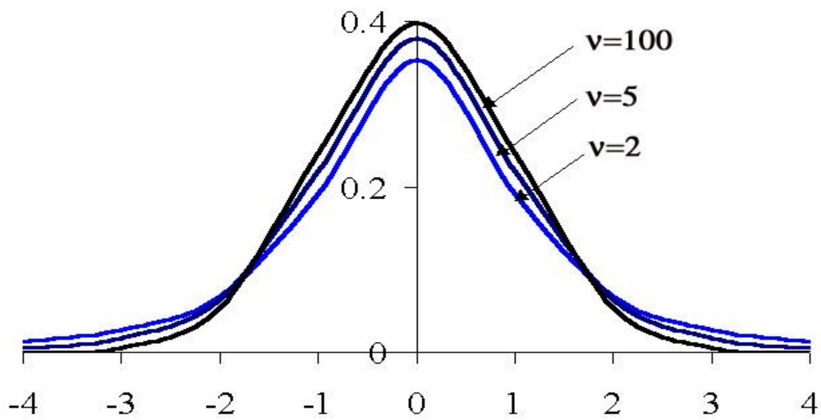
$$f(t) = \frac{\Gamma(\frac{\upsilon+1}{2})}{\Gamma(\frac{\upsilon}{2})\sqrt{\pi\upsilon}}(1+\frac{t^2}{\upsilon})^{-(\frac{\upsilon+1}{2})} \quad \text{เมื่อ } -\infty \le t < \infty \quad ....(4.13)$$

มีองศาเสรี (Degree of Freedom, υ) เท่ากับ n-1

$$\mu_t = 0 \quad .... (4.14)$$

$$\sigma^2_t = \frac{\upsilon}{(\upsilon-2)} \quad \text{เมื่อ } \upsilon > 3 \quad .... (4.15)$$

ตารางการแจกแจงค่า t แสดงอยู่ในตารางที่ A6 ของภาคผนวกท้ายเล่ม



รูปที่ 4.5 การแจกแจงของ t ที่องศาเสรีต่าง ๆ กัน

ความสัมพันธ์ระหว่างตัวแปรสุ่ม Z, T, χ^2

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$t = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}\cdot\frac{\sigma\sqrt{n-1}}{\sigma\sqrt{n-1}} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\cdot\frac{\sigma\sqrt{n-1}}{S\sqrt{n-1}}$$

$$= \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\cdot\frac{1}{\frac{S\sqrt{n-1}}{\sigma\sqrt{n-1}}} = \frac{z}{\sqrt{\frac{\chi^2}{n-1}}} = \frac{z}{\sqrt{\frac{\chi^2}{\upsilon}}} \quad ....(4.16)$$

สามารถใช้ t อธิบายการแจกแจงความน่าจะเป็นของค่าเฉลี่ยของตัวอย่าง ($\bar{X}$) กรณีไม่ทราบ σ^2 และ $n < 30$

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$ มี pdf ดังสมการที่ 4.13

ตัวอย่างที่ 4.2 จงหาค่า k ที่ทำให้

$P(k \leq T < -1.761) = 0.045$

เมื่อ n = 15 ซึ่งสุ่มจากประชากรปกติ

วิธีทำ $\upsilon = n-1 = 15-1 = 14$

$$P(k \leq T < -1.761) = P[T < -1.761] - P(T < k)$$
$$= 0.05 - P(T < k)$$
$$= 0.045$$
$$P(T < k) = 0.05 - 0.045 = 0.005$$
$$k = -2.977$$

ตัวอย่างที่ 4.3 อายุการใช้งานของหลอดไฟฟ้ามีการแจกแจงแบบปกติ มี μ = 1,410 ชั่วโมง ให้สุ่มตัวอย่างหลอดไฟฟ้า จำนวน 25 หลอด คำนวณหาค่าอายุการใช้งานเฉลี่ยและส่วนเบี่ยงเบนมาตราฐานได้เท่ากับ 1,360 และ 200 ชั่วโมงตามลำดับ จงคำนวณหาความน่าจะเป็นที่อายุการใช้งานของหลอดไฟฟ้าจะมากกว่า 1,360 ชั่วโมง

วิธีทำ ให้ $\bar{X}$ = อายุการใช้งานเฉลี่ยของหลอดไฟฟ้าเป็นชั่วโมง

เนื่องจาก $n < 30$ และ $S = 200$ ชั่วโมง

$$P(\bar{X} > 1360 \text{ ชม}) = P\left(T > \frac{1360-1410}{\frac{200}{\sqrt{25}}}\right)$$
$$= P(T > -1.25)$$

$\upsilon = n-1 = 25-1 = 24$

จากตาราง A5

$P(T_{24} \leq 1.318) = 0.9$

$P(T_{24} \leq 0.685) = 0.75$

$P(T_{24} \geq -1.318) = 0.9$

$P(T_{24} \geq -0.685) = 0.75$

$$P(T_{24} \geq -1.25) = 0.9 - \frac{(0.9-0.75)(1.318-1.25)}{(1.318-0.685)}$$

$$= 0.9 - 0.016$$

$$= 0.884$$

4.5.3 การแจกแจงของผลต่างของค่าเฉลี่ยของ 2 ตัวอย่าง ($\bar{X}_1$-$\bar{X}_2$)

(1) กรณีทราบประชากรทั้ง 2 ชุด (ทราบ σ_1^2 , σ_2^2)

ประชากรทั้ง 2 ชุดเป็นอิสระต่อกัน

ชุดที่	ประชากร	ตัวอย่าง
I	$X_1 \sim n(x_1; \mu_1, \sigma_1)$ ทราบการแจกแจง	$(x_{1,1}, x_{1,2}, \ldots\ldots x_{1,n1}) \rightarrow \bar{x}_1 \rightarrow \bar{X}_1$ $\bar{X}_1 \sim n(\bar{x}_1; \mu_1, \sigma_1/\sqrt{n_1})$
II	$X_2 \sim n(x_2; \mu_2, \sigma_2)$ ทราบการแจกแจง	$(x_{2,1}, x_{2,2}, \ldots\ldots x_{2,n2}) \rightarrow \bar{x}_2 \rightarrow \bar{X}_2$ $\bar{X}_2 \sim n(\bar{x}_2; \mu_2, \sigma_2/\sqrt{n_2})$

$$(\bar{X}_1-\bar{X}_2) \sim n((\bar{x}_1-\bar{x}_2); \mu_{\bar{x}1-\bar{x}2}, \sigma_{\bar{x}1-\bar{x}2})$$

$$\mu_{\bar{x}1-\bar{x}2} = E(\bar{X}_1-\bar{X}_2) = \mu_1-\mu_2$$

$$\sigma^2_{\bar{x}1-\bar{x}2} = Var(\bar{X}_1-\bar{X}_2) = \sigma^2_{\bar{x}1} + \sigma^2_{\bar{x}2}$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$z = \frac{(\bar{x}1 - \bar{x}2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.17)$$

ตัวอย่างที่ 4.4 สุ่มตัวอย่างประชากร 2 ชุด อย่างอิสระจากประชากรซึ่งมีการแจกแจงปกติ โดยมีข้อมูลดังต่อไปนี้

$n_1 = 5$, $n_2 = 4$

$\mu_1 = 50$, $\mu_2 = 40$

$\sigma^2_1 = 9, \qquad \sigma^2_2 = 4$

จงหา $P[(\bar{X}_1-\bar{X}_2) \leq 8.2]$

วิธีทำ

$$P[(\bar{X}_1-\bar{X}_2) \leq 8.2] = P\left(\frac{(\bar{X}1-\bar{X}) - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \leq \frac{8.2-(50-40)}{\sqrt{\frac{9}{5}+\frac{4}{4}}}\right)$$

$$= P\left(Z \leq \left(\frac{8.2-(50-40)}{\sqrt{\frac{9}{5}+\frac{4}{4}}}\right)\right)$$

$$= P\left(Z \leq \left(\frac{8.2-10}{\sqrt{\frac{36+20}{20}}}\right)\right)$$

$$= P\left(Z \leq -\frac{1.8}{\sqrt{2.8}}\right) = P(Z \leq -1.0757)$$

$$= 0.1401$$

(2) กรณีไม่ทราบประชากรทั้ง 2 ชุด (ไม่ทราบ σ^2_1, σ^2_2) แต่ทราบว่า $\sigma^2_1 = \sigma^2_2$

$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{Sp\sqrt{\frac{1}{n1}+\frac{1}{n2}}} \qquad \text{....................(4.18)}$$

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)} \qquad \text{....................(6.19)}$$

ตัวอย่างที่ 4.5 สุ่มทดสอบความแข็งโลหะ 2 ชนิด ซึ่งมีค่าเฉลี่ย μ_1= 20.5 หน่วย และ μ_2= 19.5 หน่วย สุ่มเลือกโลหะชนิดที่ 1 จำนวน 9 ชิ้น โลหะชนิดที่ 2 จำนวน 7 ชิ้น คำนวณค่าส่วนเบี่ยงเบนมาตราฐานได้ 1.2 และ 1.5 หน่วยตามลำดับ ถ้าสมมติว่าความแข็งโลหะทั้งสองชนิดมีการแจกแจงแบบปกติและค่าส่วนเบี่ยงเบนมาตราฐานเท่ากัน จงหาความน่าจะเป็นที่ค่าเฉลี่ยของความแข็งของโลหะชนิดแรกจะแข็งกว่าโลหะชนิดที่ 2 มากกว่า 1.9 หน่วย

วิธีทำ

n_1	=	9	n_2	=	7
S_1	=	1.2	S_2	=	1.5
μ_1	=	20.5	μ_2	=	19.5

โลหะทั้ง 2 มีการแจกแจงแบบปกติ มี $\sigma^2_1 = \sigma^2_2$ แต่ไม่ทราบ σ^2_1, σ^2_2

$n_1, n_2 < 30$

เลือกใช้การแจกแจงแบบ T

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$= \frac{(9-1)(1.2)^2 + (7-1)(1.5)^2}{9+7-2}$$

$$S_p^2 = 1.337\ ;\ \upsilon = 9+7-2 = 14$$

$$P(\bar{X}_1-\bar{X}_2 > 1.9) = P\left(T > \frac{1.9-(20.5-19.5)}{1.337\sqrt{\frac{1}{9}+\frac{1}{7}}}\right)$$

$$= P(T > 1.335)$$

จากตาราง A5 เมื่อ $\upsilon = 14$

$$P(T \le 1.345) = 0.9 \rightarrow P(T > 1.345) = 0.1$$

$$P(T \le 0.692) = 0.75 \rightarrow P(T \ge 0.692) = 0.25$$

$$P(T > 1.335) = 0.1 + \frac{(0.25-0.1)(1.345-1.335)}{(1.345-0.692)}$$

$$= 0.1+0.002$$

$$= 0.102$$

(3) กรณีประชากร 2 ชุด ไม่ทราบ σ^2_1, σ^2_2 แต่ทราบว่า $\sigma^2_1 \neq \sigma^2_2$

$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n1}+\frac{S_2^2}{n2}}} \quad \text{....................(4.20)}$$

$$\upsilon = \frac{\left(\frac{S_1^2}{n1}+\frac{S_2^2}{n2}\right)^2}{\left(\frac{S_1^2}{n1}\right)\left(\frac{1}{n1-1}\right)+\left(\frac{S_2^2}{n2}\right)\left(\frac{1}{n_2-1}\right)} \quad \text{....................(4.21)}$$

ตัวอย่างที่ 4.6 บริษัทผู้ผลิตอ้างว่าค่าเฉลี่ย (μ) เวลาออกฤทธิ์ของยากับคนไข้ชายและหญิง เท่ากับ 4.8 และ 4.4 นาที ตามลำดับ เพื่อทดสอบคำกล่าวอ้างดังกล่าวเป็นจริงหรือไม่ ได้ทำการทดลองใช้ยากับคนใช้ชาย 13 คน หญิง 16 คน พบว่าค่าส่วนเบี่ยงเบนมาตราฐานที่คำนวณได้จากตัวอย่างเท่ากับ 0.8 และ 0.9 นาที สำหรับชายและหญิงตามลำดับ ถ้าระยะเวลาในการออกฤทธิ์ของยาดังกล่าวมีการแจกแจงแบบปกติ ไม่ทราบค่าส่วนเบี่ยงเบนมาตราฐานที่แท้จริง แต่ทราบว่า $\sigma^2_1 \neq \sigma^2_2$ จงหาความน่าจะเป็นที่ค่าเฉลี่ยของเวลาออกฤทธิ์ของชายจะมากกว่าหญิง มากกว่า 0.94 นาที

วิธีทำ

$n_1 = 13 \qquad n_2 = 16$

$\mu_1 = 4.8$ นาที $\mu_2 = 4.4$ นาที

$S_1 = 0.8$ นาที $S_2 = 0.9$ นาที

จงหา $P[(\bar{X}_1-\bar{X}_2) > 0.94$ นาที$]$

$\sigma^2_1 \neq \sigma^2_2$ แต่ไม่ทราบค่า และ $n_1, n_2 < 30$

เลือกใช้การแจกแจงแบบ T

$$\upsilon = \frac{S_1^2/n_1 + S_2^2/n_2}{\frac{S_1^2}{n_1}\left(\frac{1}{n_1-1}\right) + \frac{S_2^2}{n_2}\left(\frac{1}{n_2-1}\right)}$$

$$= \frac{(0.8)^2/13 + (0.9)^2/16}{\left(\frac{0.8^2}{13}\right)\left(\frac{1}{13-1}\right) - \left(\frac{0.9^2}{16}\right)\left(\frac{1}{16-1}\right)}$$

$$= 27$$

$$P[(\bar{X}_1-\bar{X}_2) > 0.94] = P\left[T > \frac{0.94-(4.8-4.4)}{\sqrt{\frac{0.64}{13}+\frac{0.81}{16}}}\right]$$

$$P(T > 1.708) \approx 0.05$$

4.5.4 การแจกแจงของค่าเฉลี่ยของผลต่างของค่าที่สุ่มเป็นคู่

กรณีข้อมูลสุ่มจากประชากรชุดเดียวกันแต่สุ่มเป็นคู่ (Paired Observation)

ให้ $\bar{D}$ คือตัวแปรสุ่ม ซึ่งเท่ากับค่าเฉลี่ยของผลต่างของค่าที่สุ่มเป็นคู่ กรณีนี้ค่าที่สุ่มแบบคู่จะไม่อิสระต่อกัน

$$t = \frac{\bar{d} - \mu_{\bar{d}}}{S_{\bar{d}}} \qquad \text{..................................(4.22)}$$

เมื่อ $d = x_1 - x_2$ (ผลต่างของค่าที่สุ่มเป็นคู่)

$$\bar{d} = \frac{\sum(x_1 - x_2)}{n}$$

$$\mu_{\bar{d}} = E(\bar{D}) = E\left[\left(\frac{\sum^n D}{n}\right)\right] = \frac{1}{n}\left(\sum^n E(D)\right) = \frac{1}{n}.n.\mu_d$$

$$= \mu_d \qquad \text{..................................(4.23)}$$

$$S^2_{\bar{d}} = Var(\bar{D}) = Var\left(\frac{\sum^n D}{n}\right) = \frac{1}{n^2}(nS_d{}^2)$$

$$S_{\bar{d}} = \frac{S_d}{\sqrt{n}} \qquad \text{.................................}(4.24)$$

ตัวอย่างที่ 4.7 โครงการควบคุมอาหารแบบใหม่ สามารถลดน้ำหนักได้ 10 กก. ใน 2 สัปดาห์ ถ้าสุ่มตัวอย่างผู้หญิงที่เข้าโครงการจำนวน 7 คน วัดน้ำหนักก่อนและหลังเข้าโครงการ ปรากฏว่าค่าส่วนเบี่ยงเบนเป็นมาตราฐานเท่ากับ 5.4 กก. จงหาความน่าจะเป็นที่ค่าเฉลี่ยของน้ำหนักที่ลดลงมากกว่า 14 กก.

วิธีทำ ให้ $\bar{D}$ = ตัวแปรสุ่มแสดงค่าน้ำหนักที่ลดลงหลังเข้าโครงการควบคุมอาหาร

μ_d = 10 กก. ใน 2 สัปดาห์

$n = 7$

$S_d = 5.4$

จงหา $P(\bar{D} > 14$ กก.$)$

$$= P\left(T > \frac{14-10}{5.4/\sqrt{7}}\right)$$

$$= P(T > 1.96)$$

$$\approx 0.05$$

4.5.5 การแจกแจงของสัดส่วนของตัวอย่าง $(\hat{P})$

$$\hat{P} \sim n\left(\hat{p}\,;\, p, \sqrt{\frac{pq}{n}}\right)$$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \qquad \text{.......................................}(4.25)$$

*** ให้ X = ตัวแปรสุ่มทวินาม มี pmf เท่ากับ b(x; n,p)**

$$\mu_x = np \qquad \text{.......................................}(4.26)$$

$$\sigma^2_x = npq \qquad \text{.......................................}(4.27)$$

ให้ $\hat{p} = \frac{x}{n}$ เป็นตัวประมาณค่าของ p ซึ่งเป็นพารามิเตอร์ของการแจกแจงทวินาม

ถ้าพิจารณาว่า $\hat{P}$ เป็นตัวแปรสุ่ม

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{np}{n} = p$$

$$Var(\hat{P}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{npq}{n^2} = \frac{pq}{n}$$

ตัวอย่างที่ 4.8 จงหาความน่าจะเป็นของการโยนเหรียญ 120 ครั้ง แล้วได้หัว

(ก) ระหว่าง 40% ถึง 60%

(ข) มากกว่าหรือเท่ากับ $\frac{5}{8}$ ของจำนวนครั้งที่โยน

ในการโยนเหรียญ ความน่าจะเป็นในการออกหัว = $p = 0.5$

ความน่าจะเป็นในการออกก้อย = $q = 0.5$

วิธีทำ $P(H) = p = 0.5$

$n = 120$

$\mu_{\hat{p}} = p = 0.5$

$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.5x0.5}{120}} = 0.04566$

ก. $P(40\% \leq \hat{P} < 60\%)$

$= P(0.4 \leq \hat{P} < 0.6)$

$= P[\frac{0.4-0.5}{0.04566} \leq Z < \frac{0.6-0.5}{0.04566}]$

$= P(-2.190 < Z < 2.190)$

$= 0.9714$

ข. $P(X \geq \frac{5}{8} * 120)$

$= P(X \geq 75) = P(X' \geq 74.5)$

$= P(\hat{P} \geq \frac{74.5}{120} = 0.621)$

$= P(Z \geq \frac{0.621-0.5}{0.04566})$

$= P(Z \geq 2.651)$

$= 1-0.9960$

$= 0.0040$

4.5.6 การแจกแจงของผลต่างของสัดส่วนของ 2 ตัวอย่าง $(\hat{P}_1-\hat{P}_2)$

$$(\hat{P}_1 - \hat{P}_2) \sim n((\hat{p}_1 - \hat{p}_2); (p_1 - p_2), \sqrt{\frac{p_1q_1}{n1} + \frac{p_2q_2}{n2}})$$

$$z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n1}+\frac{p_2q_2}{n2}}} \quad \text{(ประชากร 2 ชุด)} \quad ....(4.28)$$

ตัวอย่างที่ 4.9 ผลการสำรวจความชอบแชมพูยี่ห้อหนึ่งปรากฏว่า % ชาย : % หญิง = 35 : 30 ถ้าสัมภาษณ์ชาย 500 หญิง 500 จงหาความน่าจะเป็นที่ชายชอบมากกว่าหญิงมากกว่า= 10%

วิธีทำ

ให้ $\hat{P}_1$ คือตัวแปรสุ่มแสดง % ชายที่ชอบ มี $p_1 = 0.35$

$\hat{P}_2$ คือตัวแปรสุ่มแสดง % หญิงที่ชอบ มี $p_2 = 0.30$

$$\mu_{\hat{p}_1-\hat{p}_2} = p_1-p_2 = 0.35-0.30 = 0.05$$

$$\sigma^2_{\hat{p}_1-\hat{p}_2} = \frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}$$

$$= \left(\frac{0.35\text{x}0.65}{500}+\frac{0.3\text{x}0.7}{500}\right)$$

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{0.000875} = 0.0296$$

$$P(\hat{P}_1-\hat{P}_2 > 0.1) = P[Z \geq \frac{0.1-0.05}{0.0296}]$$

$$= P[Z \geq 1.689]$$

$$= 0.0455$$

4.5.7 การแจกแจงแบบไคสแควร์ (Chi-Squared distributions)

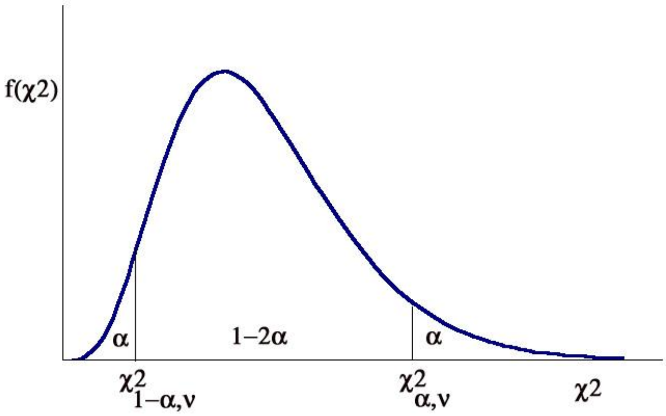
ตามที่ได้กล่าวมาแล้วในหัวข้อ 3.3.5 ของบทที่ 3 การแจกแจงแบบไคสแควร์ คือการแจกแจงแบบแกมม่า ซึ่งมี $\alpha = \frac{\upsilon}{2}, \beta = 2$

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2} \quad \text{.................................(4.29)}$$

มี pdf ดังสมการที่ 3.40

มีองศาเสรี (υ) = n-1

ให้ $\chi^2_{\alpha,\upsilon}$ คือค่า χ^2 ที่ความน่าจะเป็นแบบมากกว่า (Exceedence Probability) เท่ากับ α และองศาเสรีเท่ากับ υ ดังรูปที่ 4.6



รูปที่ 4.6 กราฟการแจกแจงไคสแควร์

คุณสมบัติที่สำคัญของการแจกแจง χ^2 ที่ควรทราบคือ (ดูรูปที่ 4.10 ประกอบ)

$P(\chi^2 \geq \chi^2_{\alpha,\upsilon}) = \alpha$ ....(4.30)

$P(\chi^2 < \chi^2_{\alpha,\upsilon}) = 1-\alpha$

$= 1-P(\chi^2 \geq \chi^2_{\alpha,\upsilon})$

$= P(\chi^2 > \chi^2_{1-\alpha,\upsilon})$ ....(4.31)

$P(\chi^2 < \chi^2_{1-\alpha,\upsilon}) = \alpha$

$= P(\chi^2 \geq \chi^2_{\alpha,\upsilon})$ ....(4.32)

ตัวอย่างที่ 4.10 จงหาค่า χ^2 ที่ความน่าจะเป็นต่าง ๆ ดังนี้

(1) $\chi^2_{0.01}$ เมื่อ $\upsilon = 18$

(2) $\chi^2_{0.99}$ เมื่อ $\upsilon = 18$

(3) χ_α^2 ที่ทำให้ $P(\chi^2 < \chi_\alpha^2) = 0.99$ เมื่อ $\upsilon = 4$

(4) $P(\chi^2_{1-\alpha} < \chi^2 < \chi_\alpha^2) = 0.95$ เมื่อ $\upsilon = 15$

วิธีทำ $P(\chi^2 \geq \chi^2_{0.01,\,18}) = 0.01$

จากตารางที่ A4

$\chi^2_{0.01,\,18} = 34.805$

(2) $\chi^2_{0.99,\,18} = 7.015$

(3) $P(\chi^2 < \chi^2_{\alpha,\,4}) = 0.99$

$P(\chi^2 \geq \chi^2_{\alpha,\,4}) = 1-.099 = 0.01$

$\chi^2_{\alpha,\,4} = \chi^2_{0.01,\,4} = 13.277$

(4) $P(\chi^2_{1-\alpha,15} < \chi^2 < \chi^2_{\alpha,15}) = 0.95 = (1-2\alpha)$

$P(\chi^2 < \chi^2_{\alpha,15}) - P(\chi^2 < \chi^2_{1-\alpha,15}) = 0.95$

$1 - P(\chi^2 > \chi^2_{\alpha,15}) - P(\chi^2 > \chi^2_{\alpha,15}) = 0.95$

$1 - 2\,P(\chi^2 > \chi^2_{\alpha,15}) = 0.95$

$P(\chi^2 > \chi^2_{\alpha,15}) = 0.025$

$\chi^2_{\alpha,15} = \chi^2_{0.025,15} = 27.488$

$\chi^2_{1-\alpha,15} = \chi^2_{0.975,15} = 6.262$

ตัวอย่างที่ 4.11 สุ่มตัวอย่างจำนวน 12 ตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ มี $\sigma^2 = 100$ จงหาความน่าจะเป็นที่จะสุ่มตัวอย่างได้ S มากกว่า 6.45

วิธีทำ $P(S > 6.45)$

$$= P\left[\frac{(n-1)(S)^2}{\sigma^2} > \frac{(12-1)(6.45)^2}{100}\right]$$

$$= P(\chi^2 > 4.576) = \alpha$$

องศาเสรี $= \upsilon = n-1 = 12-1 = 11$

$\therefore$ $P(\chi^2 > 4.576) = \alpha = 0.95$

4.5.8 การแจกแจงแบบ F (F-distributions)

$$F = \frac{\chi_1^2/\upsilon 1}{\chi_2^2/\upsilon 2} = \frac{\dfrac{(n_1-1)}{(n_1-1)}\dfrac{S_1^2}{\sigma_1^2}}{\dfrac{(n_2-1)}{(n_2-1)}\dfrac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \quad \text{...........(4.33)}$$

มีองศาเสรี $\upsilon 1$, $\upsilon 2$

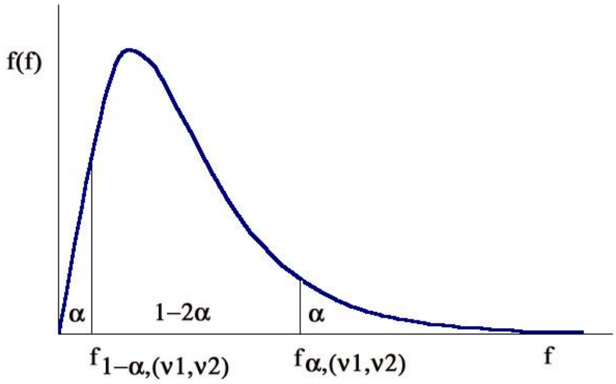
$\upsilon 1 = n_1 - 1$

$\upsilon 2 = n_2 - 1$

$$h(f) = \left[\frac{\Gamma\left(\frac{\upsilon_1+\upsilon_2}{2}\right)\left(\frac{\upsilon_1}{\upsilon_2}\right)^{\upsilon./2} f^{\frac{\upsilon_1}{2}-1}}{\Gamma(\upsilon 1/2)\Gamma(\upsilon 2/2)\left\{1+\frac{\upsilon_1}{\upsilon_2}.f\right\}^{\frac{\upsilon_1+\upsilon_2}{2}}}\right] \quad \text{......................(4.34)}$$

เมื่อ $0 \leq f < \infty$

กราฟการแจกแจงค่า F มีลักษณะดังรูปที่ 4.7 ตารางค่า f มีค่าดังตารางที่ A7 ของภาคผนวก



รูปที่ 4.7 กราฟการแจกแจงแบบ F

คุณสมบัติที่สำคัญของการแจกแจงของ F ได้แก่ (ดูรูปที่ 4.11 ประกอบ)

$P[F > f_{\alpha,(\upsilon1,\upsilon2)}] = \alpha$ ....(4.35)

$P(F < f_{\alpha,(\upsilon1,\upsilon2)}] = 1-\alpha$ ....(4.36)

$P[F < f_{\alpha,(\upsilon1,\upsilon2)}] = 1-P[F > f_{\alpha,(\upsilon1,\upsilon2)}]$ ....(4.37)

$P(F > f_{1-\alpha,(\upsilon1,\upsilon2)}] = 1-\alpha$

$= 1-P[F > f_{\alpha,(\upsilon1,\upsilon2)}]$ ....(4.38)

เนื่องจากตารางค่า f มีเฉพาะค่า f_α เมื่อ $\alpha = 0.01$ และ 0.05 เท่านั้น ค่า $f_{1-\alpha}$ จะหาได้จากสมการ

$$f_{1-\alpha,(\upsilon1,\upsilon2)} = \frac{1}{f_{\alpha,(\upsilon_2,\upsilon_1)}} \qquad ....(4.39)$$

เช่น $f_{0.95,(\upsilon1,\upsilon2)} = \dfrac{1}{f_{0.05(\upsilon_2,\upsilon_1)}}$

$$\mu = \frac{\upsilon_2}{\upsilon_2 - 2} \quad \text{เมื่อ } \upsilon_2 > 2 \qquad ....(4.40)$$

$$\sigma^2 = \frac{2\upsilon_2^2(\upsilon_1 + \upsilon_2 - 2)}{(\upsilon_2 - 4)\,\upsilon_1(\upsilon_2 - 2)^2} \quad \text{เมื่อ } \upsilon_2 > 4 \qquad .....(4.41)$$

ตัวอย่างที่ 4.12 จงหาค่า f และความน่าจะเป็นจากกรณีต่าง ๆ ดังต่อไปนี้

(1) $f_{0.05}$ เมื่อ $\upsilon_1 = 7, \upsilon_2 = 15$

(2) $f_{0.95}$ เมื่อ $\upsilon_1 = 7, \upsilon_2 = 15$

(3) จงหาค่า f ถ้า $P(f_{1-\alpha} < F < f_\alpha) = 0.9$ เมื่อ $\upsilon_1 = 20$ และ $\upsilon_2 = 30$

(4) $P(\frac{1}{2.62} < F < 2.48)$ เมื่อ $\upsilon_1 = 12, \upsilon_2 = 15$

วิธีทำ

(1) $f_{0.05,(7,15)} = 2.71$

(2) $f_{0.95,(7,15)} = \frac{1}{f_{0.05,(15,7)}} = \frac{1}{3.51} = 0.2849$

(3)
$$
\begin{aligned}
P[f_{1-\alpha,(20,30)} < F < f_{\alpha,(20,30)}] &= 0.9 \quad (=1-2\alpha)\\
P[F < f_{\alpha,(20,30)}] - P[F < f_{1-\alpha,(20,30)}] &= 0.9\\
1-P[F > f_{\alpha,(20,30)}] - P[F > f_{\alpha,(20,30)}] &= 0.9\\
1-2P[F > f_{\alpha,(20,30)}] &= 0.9\\
P[F > f_{\alpha,(20,30)}] &= 0.05\\
\alpha &= 0.05\\
f_{1-\alpha,(20,30)} &= f_{0.95,(20,30)}\\
&= \frac{1}{f_{0.05,(30,20)}}\\
&= \frac{1}{2.04} = 0.492\\
f_{\alpha,(20,30)} &= f_{0.05,(20,30)}\\
&= 1.93
\end{aligned}
$$

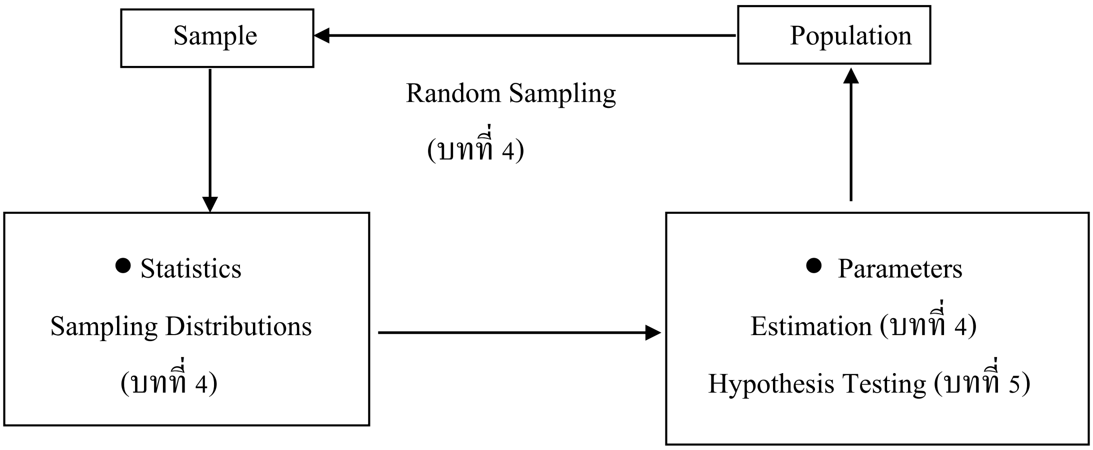
(4) $P(\frac{1}{2.62} < F < 2.48)$
$$
\begin{aligned}
&= P[\frac{1}{f_{0.05,(15,12)}} < F < f_{0.05,(12,15)}]\\
&= P[f_{0.95,(12,15)} < F < f_{0.05,(12,15)}]\\
&= 0.95 - 0.05 = 0.9
\end{aligned}
$$

4.6 สรุปการแจกแจงของตัวอย่างสุ่ม

การแจกแจงของตัวอย่างสุ่มที่กล่าวถึงในบทนี้ สามารถสรุปได้ดังตารางที่ 4.1

4.7 แนวความคิดในการประมาณค่า (Estimation Concept)

ในหัวข้อที่แล้วได้กล่าวถึงเทคนิคของการสุ่มตัวอย่าง การหาค่าสถิติที่สำคัญและการแจกแจงของตัวอย่างสุ่ม ต่อไปจะได้กล่าวถึงการประมาณค่าพารามิเตอร์ และการหาขนาดตัวอย่างที่เหมาะสม และในบทที่ 5 จะได้กล่าวถึงการทดสอบสมมติฐานของพารามิเตอร์ของประชากร (Hypothesis Testing) ซึ่งอาศัยพื้นฐานจากทฤษฎีการแจกแจงของตัวอย่างสุ่ม (Sampling Distributions) ที่กล่าวมาแล้ว โดยการประมาณค่าพารามิเตอร์จะเน้นการตอบคำถามที่ว่า "***พารามิเตอร์ของประกรมีค่าเท่าใด***" และการทดสอบสมมติฐานในบทที่ 5 จะเน้นการตอบคำถามที่ว่า "***ประชากรมีค่าพารามิเตอร์เท่านั้นจริงหรือไม่***" ความสัมพันธ์ของเนื้อหาในบทที่ 4 และ 5 แสดงอยู่ในรูปที่ 4.8



รูปที่ 4.8 แนวความคิดเกี่ยวกับการสุ่มตัวอย่าง การประมาณค่าและการทดสอบสมมติฐานของพารามิเตอร์ของประชากร

หลักการ วัตถุประสงค์ และวิธีการในการประมาณค่าพารามิเตอร์ของประชากรที่จะกล่าวถึงในหัวข้อต่อไปนี้สามารถนำไปเปรียบเทียบกับการสุ่มตัวอย่างในหัวข้อที่ 4.3 ถึง 4.6 ได้ดังตารางที่ 4.2

4.8 วิธีการประมาณค่า

วิธีการประมาณค่ามี 2 วิธี คือ

- การประมาณค่าเป็นจุด หรือค่าเดียว (Point Estimate)
- การประมาณช่วงค่า (Interval Estimate)

4.8.1 การประมาณค่าเป็นจุด หรือค่าเดี่ยว (Point Estimate)

ตัวอย่างตัวประมาณค่า (Estimator)ที่สำคัญได้แก่

$\bar{x}$ เป็นตัวประมาณค่า (Estimator) ของ μ

s^2 เป็นตัวประมาณค่า (Estimator) ของ σ^2

$\hat{p}$ เป็นตัวประมาณค่า (Estimator) ของ p

- **ตัวประมาณค่าที่ไม่ลำเอียง (Unbiased Estimator)**

ถ้า Expected Value ของตัวประมาณค่ามีค่าเท่ากับค่าของพารามิเตอร์ที่ต้องการประมาณ แสดงว่าตัวประมาณค่าตัวนั้นเป็นตัวประมาณค่าแบบไม่ลำเอียง ตามที่เคยกล่าวไว้ในหัวข้อ 6.4.2 เช่น

$$E(\bar{X}) = \mu$$

$$E\left(S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}\right) = \sigma^2$$

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{np}{n} = p$$

- **ถ้า $\hat{\theta}$ เป็น Unbiased Estimator ของ θ และ**

$$Var(\hat{\theta}) = \frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]} \qquad \text{.......................................(4.42)}$$

= มีค่าน้อยที่สุด

แสดงว่าตัวประมาณค่าที่ไม่ลำเอียงมีความค่าความแปรปรวนน้อยที่สุด

ถ้า $\hat{\theta}_1$, $\hat{\theta}_2$ เป็นตัวประมาณค่าที่ไม่ลำเอียง ของ θ และ

ถ้า $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

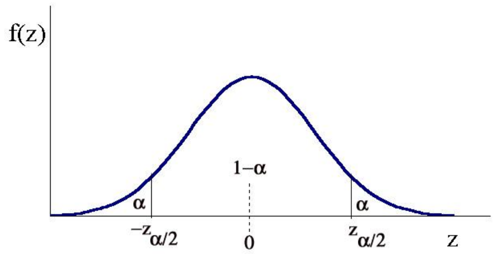
$\hat{\theta}_1$ จะเป็นตัวประมาณค่าที่ไม่ลำเอียงที่ดีกว่า $\hat{\theta}_2$ ซึ่งสามารถวัดได้โดยใช้

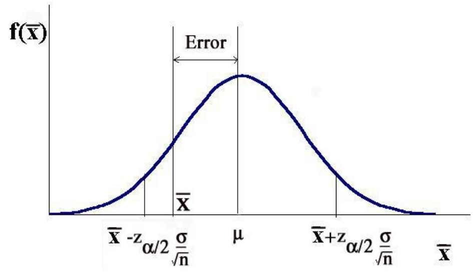
$$\text{(RER) Relative Efficiency Ratio} = \frac{Var(\theta_2)}{Var(\theta_1)} \qquad \text{...........................(4.43)}$$

ถ้า RER < 1 แสดงว่า $\hat{\theta}_2$ ดีกว่า $\hat{\theta}_1$

RER > 1 แสดงว่า $\hat{\theta}_1$ ดีกว่า $\hat{\theta}_2$

ตารางที่ 4.2 เปรียบเทียบเนื้อของการสุ่มตัวอย่างในบทที่ 6 และการประมาณค่าในบทที่ 7

<table>
<tr><th></th><th>บทที่ 6 การสุ่มตัวอย่าง</th><th>บทที่ 7 การประมาณค่า</th></tr>
<tr><td>หลักการ</td><td>$\bar{X}$,$(\bar{X}_1-\bar{X}_2)$, $\hat{P}$,$(\hat{P}_1-\hat{P}_2)$,
$\bar{D}$, S^2 , $S_1^2\,/\,S_2^2$ เป็นตัวแปรสุ่ม
(Random Variable) มี pdf = f(x)</td><td>$\bar{x}\sim\mu$, $(\bar{x}_1-\bar{x}_2)\sim(\mu_1-\mu_2)$
$\hat{p}\sim p$, $(\hat{p}_1-\hat{p}_2)\sim(p_1-p_2)$
$\bar{d}\sim\mu_{\bar{D}}$, $s^2\sim\sigma^2$, $\frac{s_1^2}{s_2^2}\sim\frac{\sigma_1^2}{\sigma_2^2}$</td></tr>
<tr><td>วัตถุประสงค์</td><td>รู้ค่าพารามิเตอร์ตัวที่ต้องการประมาณ
เช่น μ, σ^2 ต้องการหา
(1) $P(\bar{X}\geq a) = ?$
$P(\bar{X}\leq b) = ?$
$P(a\leq\bar{X}<b) = ?$
(2) กำหนด Probability
หาค่า a, b ของ $\bar{X}$</td><td>ไม่รู้ค่าพารามิเตอร์ต้องการหา
(1) หาช่วงความเชื่อมั่น $(1-\alpha)$ ของพารามิเตอร์จากตัวประมาณค่า
(2) หาขนาดตัวอย่าง (ค่า n) ซึ่ง Error ในการประมาณค่าพารามิเตอร์ อยู่ในเกณฑ์ที่ยอมรับได้ emax</td></tr>
<tr><td>วิธีการ</td><td>$\bar{x}$, $(\bar{x}_1-\bar{x}_2)$ $\rightarrow$ Z, T
$\hat{p}$, $(\hat{p}_1-\hat{p}_2)$ $\rightarrow$ Z
$\bar{d}$ $\rightarrow$ T
S^2 $\rightarrow$ χ^2
$\frac{S_1^2}{S_2^2}$ $\rightarrow$ F</td><td>

$\bar{x}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\leq\mu<\bar{x}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$


Error= e $=|\bar{x}-\mu|=z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$
ถ้า $n=[\frac{z_{\frac{\alpha}{2}}\sigma}{e}]^2$ ค่า Error $\leq z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$</td></tr>
</table>

- **ตัวอย่างเช่นถ้าสุ่มตัวอย่างขนาด (2n+1)**

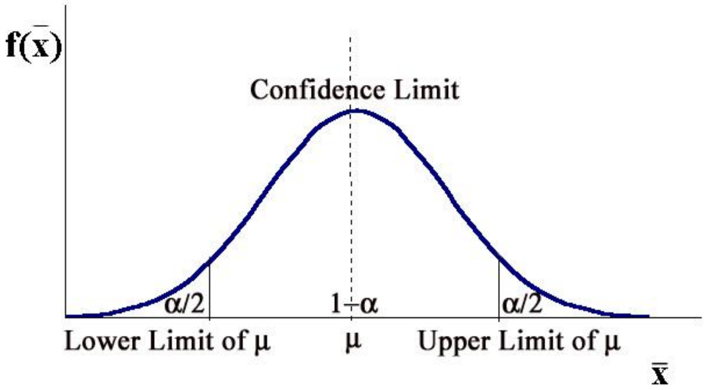
$$\text{RER} = \frac{\text{Var (Mean}, \bar{x})}{\text{Var (Median}, \tilde{x})} = \frac{\frac{\sigma^2}{(2n+1)}}{\frac{\pi\sigma^2}{4n}} = \frac{4n}{\pi(2n+1)}$$

ถ้า $n \to \infty$

$$\text{RER} = \frac{2}{\pi} < 1 \quad \text{แสดงว่า } \bar{x} \text{ ดีกว่า } \tilde{x}$$

4.8.2 การประมาณช่วงค่า (Interval Estimate)

วิธีการประมาณช่วงค่าของพารามิเตอร์ เช่น μ แสดงอยู่ในรูปที่ 4.9



รูปที่ 4.9 การประมาณช่วงค่าของ μ

ให้ $1-\alpha$= ระดับความเชื่อมั่น (Confidence Level) = ความน่าจะเป็นที่ที่พารามิเตอร์จะมีค่าอยู่ระหว่างช่วงค่าใด ๆ

$1-\alpha$ = P (LL < Parameter < UL) [2 ทาง] ....(4.44)

$1-\alpha$ = P ($-\infty$ < Parameter < UL) [ทางเดียวด้านบน] ....(4.45)

$1-\alpha$ = P (LL < Parameter < ∞) [ทางเดียวด้านล่าง] ....(4.46)

- **ช่วงความเชื่อมั่นของการแจกแจงแบบปกติ**

Confidence Level $(1-\alpha)$ 100 %	Confidence Limits z
90	$\pm$ 1.645
95	$\pm$ 1.960
98	$\pm$ 2.325
99	$\pm$ 2.575

4.9 ช่วงความเชื่อมั่นของค่าเฉลี่ย (Confidence Interval for Mean, μ)

เมื่อพิจารณาว่าค่าเฉลี่ยของตัวอย่าง ($\overline{X}$) เป็นตัวแปรสุ่ม ซึ่งค่า $\overline{x}$ จะแตกต่างกันไปตามตัวอย่างขนาด n ที่สุ่มได้ $\overline{X}$ จะมีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ $\frac{\sigma^2}{n}$ ถ้าทราบการแจกแจงของ $\overline{X}$ และกำหนดระดับความเชื่อมั่น (1-α) ซึ่งก็คือค่าความน่าจะเป็นที่ $\overline{X}$ จะมีค่าอยู่ระหว่าง a และ b นั่นเอง

$$P(a \leq \overline{X} < b) \quad = \quad 1-\alpha \qquad \text{.........................(4.47)}$$

เมื่อ α = ระดับนัยสำคัญ (Significant Level)

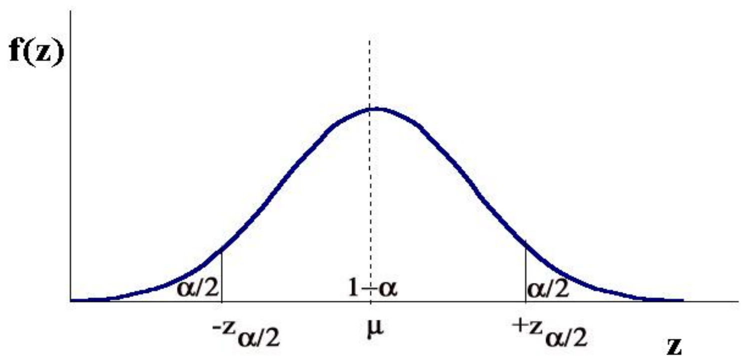
จากสมการที่ 7.3-7.5 จะสามารถนำไปวิเคราะห์หาช่วงความเชื่อมั่น (Confident Interval) ของค่าพารามิเตอร์ (μ) ของประชากรได้

4.9.1 กรณีทราบ σ^2

ประมาณช่วงความเชื่อมั่นของ μ โดยใช้การแจกแจงตัวอย่างสุ่มแบบ Z ดังรูปที่ 4.10

$$\overline{X} \sim n\,(\overline{x}\,;\mu, \frac{\sigma}{\sqrt{n}})$$

$$z \quad = \quad \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim n\ \ (z\,;0,1)$$



รูปที่ 4.10 การใช้ z ประมาณช่วงความเชื่อมั่น (1-α)ของ μ

จากรูปการแจกแจงของ z จะเขียนเป็นสมการได้ว่า

$$P(-z_{\frac{\alpha}{2}} \le Z < z_{\frac{\alpha}{2}}) = 1-\alpha \qquad \text{.............................(4.48)}$$

(1-α) คือความน่าจะเป็นที่ Z จะมีค่าอยู่ระหว่างขีดจำกัดล่าง (Lower Confidence Limit, $-z_{\frac{\alpha}{Z}}$) และขีดจำกัดบน (Upper Confidence Limit, $z_{\frac{\alpha}{Z}}$)

$$P(-z_{\frac{\alpha}{2}} \le \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}) = 1-\alpha$$

$$P(-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \bar{X} - \mu < z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(\bar{X}+z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} > \mu \ge \bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(\bar{X}-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha \qquad \text{.............................(4.49)}$$

ที่ระดับความเชื่อมั่น (1-α) จะสามารถหาช่วงความเชื่อมั่นของ μ (หรือเรียกสั้น ๆ ว่า ช่วงความเชื่อมั่น (1-α) ของ μ) ได้จากสมการ

$$\bar{x}-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu < \bar{x}+z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{[2 ทาง]} \qquad \text{.............................(4.50)}$$

$$-\infty \le \mu < \bar{x}+z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{[ทางเดียวด้านบน]} \qquad \text{.............................(4.51)}$$

$$\bar{x}-z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \le \mu < \infty \quad \text{[ทางเดียวด้านล่าง]} \qquad \text{.............................(4.52)}$$

ตัวอย่างที่ 4.13 สุ่มตัวอย่างลูกหนี้ของบริษัทแห่งหนึ่งจำนวน 100 ราย ปรากฏว่า มีหนี้สินเฉลี่ย ($\bar{x}$) รายละ 2,763 บาท และมีค่าส่วนเบี่ยงเบนมาตรฐาน (σ) 674 บาท จงหาช่วงความเชื่อมั่นของหนี้สินเฉลี่ยของลูกหนี้ของบริษัท (μ) ถ้ากำหนดระดับความเชื่อมั่น (1-α) เท่ากับ 0.95

วิธีทำ กรณีนี้ทราบค่า σ จะประมาณช่วงความเชื่อมั่นของ μ แบบ 2 ทาง โดยใช้ Z

$$P(\bar{X}-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu < \bar{X}+z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 0.95$$

ดังนั้นช่วงความเชื่อมั่นที่ 0.95 ของ μ

$$\bar{x}-z_{0.025} \frac{\sigma}{\sqrt{n}} \le \mu < \bar{x}+z_{0.025} \frac{\sigma}{\sqrt{n}}$$

$$2{,}763 - 1.96 \times \frac{674}{10} \le \mu < 2{,}763 + 1.96 \times \frac{674}{10}$$

$$2{,}630.90 \le \mu < 2{,}895.10$$

สรุป หนี้สินเฉลี่ยของลูกหนี้ของบริษัทจะมีค่าอยู่ระหว่าง 2,630.90 บาท ถึง2,895.10 บาทต่อราย ด้วยระดับความเชื่อมั่นเท่ากับ 0.95

4.9.2 กรณีไม่ทราบ σ^2 แต่ n>30

ประมาณช่วงความเชื่อมั่นของ μ โดยใช้ Z

และใช้ s^2 แทน σ^2

ช่วงความเชื่อมั่น (1-α) ของ μ จะหาได้จากสมการ

$$\bar{x}-z_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}} \leq \mu < \bar{x}+z_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}} \quad \text{[2 ทาง]} \quad ....(4.53)$$

$$-\infty \leq \mu < \bar{x}+z_{\alpha}\cdot\frac{s}{\sqrt{n}} \quad \text{[ทางเดียวด้านบน]} \quad ....(4.54)$$

$$\bar{x}-z_{\alpha}\cdot\frac{s}{\sqrt{n}} \leq \mu < \infty \quad \text{[ทางเดียวด้านล่าง]} \quad ....(4.55)$$

4.9.3 กรณีไม่ทราบ σ^2 แต่ n<30

ประมาณช่วงความเชื่อมั่นของ μ โดยใช้ T

ช่วงความเชื่อมั่น (1-α) ของ μ จะหาได้จากสมการ

$$\bar{x}-t_{\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}} \leq \mu < \bar{x}+t_{\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}} \quad \text{[2 ทาง]} \quad ....(4.56)$$

$$-\infty \leq \mu < \bar{x}+t_{\alpha}\cdot\frac{S}{\sqrt{n}} \quad \text{[ทางเดียวด้านบน]} \quad ....(4.57)$$

$$\bar{x}-t_{\alpha}\cdot\frac{S}{\sqrt{n}} \leq \mu < \infty \quad \text{[ทางเดียวด้านล่าง]} \quad ....(4.58)$$

ตัวอย่างที่ 4.14 สุ่มตัวอย่างลิกไนต์จำนวน 8 ตัวอย่าง วิเคราะห์หา % ของกำมะถันได้ดังนี้ 1.01, 0.97, 1.03, 1.04, 0.98, 0.99, 1.01 และ 1.03 จงหาช่วงความเชื่อมั่น 99% ของค่าเฉลี่ย % กำมะถัน (μ) ในแร่ลิกไนต์จากแหล่งแร่นี้

วิธีทำ

$\bar{x}$ = 1.0075, s = 0.0255

υ = n-1 = 8-1 = 7

กรณีไม่ทราบ σ และ n < 30 ประมาณช่วงความเชื่อมั่นของ μ แบบ 2 ทางโดยใช้ T

ช่วงความเชื่อมั่น 99% ของค่าเฉลี่ย % ของกำมะถัน (μ)

$$\bar{x}-t_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}} \leq \mu < \bar{x}+t_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}}$$

$$1.0075 - 3.499 \times \frac{0.0255}{\sqrt{8}} \le \mu < 1.0075 + 3.499 \times \frac{0.0255}{\sqrt{8}}$$

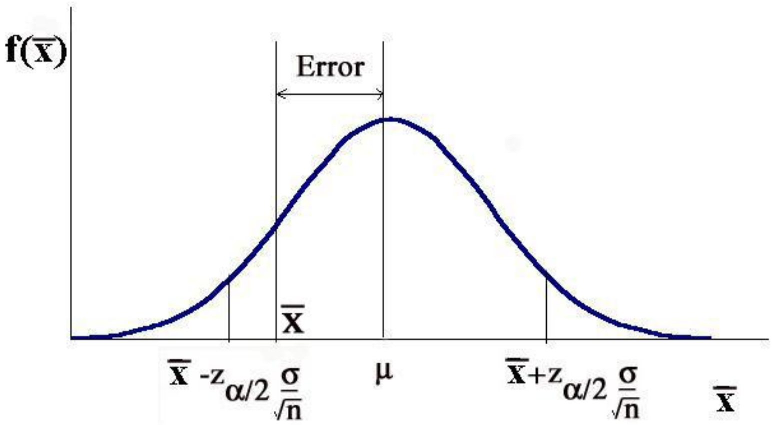
$$0.9760 \le \mu < 1.0390$$

สรุป สามารถเชื่อมั่นได้ 99% ว่าค่าเฉลี่ย % กำมะถันในแร่ลิกไนต์จากแหล่งแร่นี้มีค่าอยู่ระหว่าง 0.9760% ถึง 1.0390%

4.9.4 การหาขนาดตัวอย่างที่เหมาะสม (Appropriate Sample Size)

ในการสุ่มตัวอย่างจากประชากรที่มีค่าเฉลี่ยเท่ากับ μ สมมติได้ค่าเฉลี่ยของตัวอย่างเท่ากับ $\bar{x}$ ความผิดพลาดในการประมาณค่า μ ด้วย $\bar{x}$ คือค่าความผิดพลาด (Error) ดังแสดงในรูปที่ 4.11

$$\text{Error} = |\bar{x} - \mu| \qquad \text{.......................(4.59)}$$



รูปที่ 4.11 ความผิดพลาดในการประมาณ μ ด้วย $\bar{x}$

ในการสุ่มตัวอย่างถ้ากำหนดว่าค่าความผิดพลาดต้องไม่เกิน $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ ที่ระดับความเชื่อมั่น (1-α)

$$\text{Error} = e = |\bar{x} - \mu| = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \qquad \text{.......................(4.60)}$$

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{e} \right)^2 \qquad \text{.......................(4.61)}$$

เมื่อ n = ขนาดตัวอย่างที่ทำให้ค่าความผิดพลาดไม่เกิน e

ค่าความผิดพลาดจะแปรผันผกผันกับขนาดตัวอย่าง (n) ดังนั้นถ้าต้องการสุ่มตัวอย่างโดยให้มีค่าความผิดพลาดน้อย จะต้องสุ่มตัวอย่างจำนวนมากขึ้น (n มาก) ***ปกติก่อนการสุ่มตัวอย่างจะมีการกำหนดระดับค่าความผิดพลาดไม่ให้เกิน (e) ที่ยอมรับได้ กำหนดระดับนัยสำคัญ (α) แล้วคำนวณหาขนาดตัวอย่าง (n) จากสมการ (7.20)***

ตัวอย่างที่ 4.15 ในการประมาณค่าเฉลี่ย (μ) ของการแจกแจงปกติ ซึ่งทราบค่าส่วนเบี่ยงเบนมาตรฐาน (σ) ต้องใช้ขนาดของตัวอย่าง (n) อย่างน้อยเท่าใด จึงจะทำให้ช่วงความเชื่อมั่น 99% ของ μ มีความยาวไม่เกิน L

วิธีทำ

$$\text{Error} \leq \frac{L}{2}$$

$$n \geq \left(\frac{z_{\frac{\alpha}{2}}\sigma}{e}\right)^2$$

$$\geq \frac{z^2_{0.005}\sigma^2}{(\frac{L}{2})^2}$$

$$\geq (2.576)^2 \times 4 \times (\frac{\sigma}{L})^2$$

จำนวนตัวอย่าง (n) $\geq 26.5431\ (\frac{\sigma}{L})^2$ ตัวอย่าง

ตัวอย่างที่ 4.16 ถ้าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) เท่ากับ 0.05 จะต้องใช้ตัวอย่างขนาดเท่าใด เพื่อให้มีความเชื่อมั่น 95% ว่าความผิดพลาดของค่าประมาณของค่าเฉลี่ยจะไม่เกิน 0.01

วิธีทำ

$$n \geq \left(\frac{z_{\frac{\alpha}{2}}\sigma}{e}\right)^2$$

$$\geq \left(\frac{z_{0.025} \times 0.05}{0.01}\right)^2$$

$$\geq \left(\frac{1.96 \times 0.05}{0.01}\right)^2$$

จะต้องใช้ตัวอย่างประมาณ 97 ตัวอย่าง

4.10 ช่วงความเชื่อมั่นของผลต่างของค่าเฉลี่ย (μ_1-μ_2)

กรณีที่ $\bar{X}_1$ และ $\bar{X}_2$ เป็นตัวแปรสุ่มซึ่งค่า $\bar{x}_1$ และ $\bar{x}_2$ คือค่าเฉลี่ยของตัวอย่างขนาด n_1 และ n_2 ที่สุ่มได้ ซึ่งจะแตกต่างกันไปในการสุ่มตัวอย่างแต่ละครั้ง

4.10.1 ถ้าทราบ σ_1^2 และ σ_2^2

ประมาณช่วงความเชื่อมั่น (1-α) ของ (μ_1-μ_2) โดยใช้ Z

กรณี 2 ทาง

$$(\bar{x}_1-\bar{x}_2)-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2)+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \qquad \text{.............(4.62)}$$

ตัวอย่างที่ 4.17 สุ่มตัวอย่างหลอดไฟที่ผลิตโดยโรงงาน A 100 หลอด และโรงงาน B 50 หลอด ตรวจสอบอายุการใช้งานเฉลี่ยของหลอดไฟจากโรงงานทั้งสองได้ ($\bar{x}_1$) 1,100 ชั่วโมง และ ($\bar{x}_2$) 1,000 ชั่วโมงตามลำดับ ถ้าอายุของหลอดไฟที่ผลิตโดยโรงงาน A และโรงงาน B มีการแจกแจงปกติด้วยค่าความแปรปรวน (σ^2_1) 200 ชั่วโมง^2 และ (σ^2_2) 100 ชั่วโมง^2 ตามลำดับ จงประมาณค่าผลต่างของอายุเฉลี่ยของหลอดไฟจากโรงงานทั้งสองที่ระดับความเชื่อมั่น 95%

วิธีทำ ประมาณค่าผลต่างของอายุเฉลี่ยของหลอดไฟจาก 2 โรงงานที่สุ่มได้

$(\bar{x}_1-\bar{x}_2)$ = 1,100-1,000 = 100 ชั่วโมง

ทราบ σ^2_1 และ σ^2_2 ใช้ Z แบบ 2 ทาง

ช่วงความเชื่อมั่น 95% ของ μ_1-μ_2 คือ

$$(\bar{x}_1-\bar{x}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$(1{,}100-1{,}000)-1.96\sqrt{\frac{200}{100}+\frac{100}{50}} \le \mu_1-\mu_2 < (1{,}100-1{,}000)+1.96\sqrt{\frac{200}{100}+\frac{100}{50}}$$

$$96.08 \le \mu_1-\mu_2 < 103.92$$

สรุป ผลต่างของอายุของหลอดไฟจากโรงงาน A และโรงงาน B อยู่ระหว่าง 96.08 ถึง 103.92 ชั่วโมง ด้วยระดับความเชื่อมั่น 95%

4.10.2 ถ้าไม่ทราบ σ_1^2, σ_2^2 แต่ $n_1, n_2 > 30$

ประมาณช่วงความเชื่อมั่น (1-α) ของ (μ_1-μ_2) โดยใช้ Z และใช้ S_1^2 และ S_2^2 แทน σ_1^2 และ σ_2^2 ตามลำดับ

กรณี 2 ทาง

$$(\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{.......................(4.63)}$$

ตัวอย่างที่ 4.18 ในการตรวจสอบความเหนียวของเส้นด้าย 2 ชนิด (A และ B) ได้มีการสุ่มตัวอย่างเส้นด้ายชนิดละ 50 เส้น นำมาทดสอบโดยการดึงในสภาพเหมือนกัน ปรากฏว่า A ทนแรงดึงเฉลี่ย 78.3 กิโลกรัม ด้วยส่วนเบี่ยงเบนมาตรฐาน 5.6 กิโลกรัม และ B ทนแรงดึงเฉลี่ย 87.2 กิโลกรัม ด้วยส่วนเบี่ยงเบนมาตรฐาน 6.3 กิโลกรัม จงหาช่วงความเชื่อมั่นที่ 95% ของผลต่างของค่าเฉลี่ยแรงดึงสูงสุดของเส้นด้าย

วิธีทำ ไม่ทราบ σ_1^2 และ σ_2^2 แต่ n_1 และ $n_2 > 30$ ใช้ s_1 และ s_2 แทน σ_1 และ σ_2 ของประชากร และหาช่วงความเชื่อมั่น 95% ของ (μ_1-μ_2) โดยใช้ Z แบบ 2 ทาง

ช่วงความเชื่อมั่น 95% ของ μ_1-μ_2 คือ

$$(\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(87.2 - 78.3) - 1.96 . \sqrt{\frac{6.3^2}{50} + \frac{5.6^2}{50}} \le \mu_1 - \mu_2 < (8.72 - 78.3) + 1.96 \sqrt{\frac{6.3^2}{50} + \frac{5.6^2}{50}}$$

$$6.5636 \le \mu_1 - \mu_2 < 11.2364$$

สรุป ผลต่างค่าเฉลี่ยแรงดึงในเส้นด้าย B และ A อยู่ระหว่าง 6.56 กิโลกรัม ถึง 11.24 กิโลกรัม ที่ระดับความเชื่อมั่น 95%

4.10.3 ถ้าไม่ทราบ σ_1^2, σ_2^2 แต่ $n_1, n_2 < 30$ และ $\sigma_1^2 = \sigma_2^2 = \sigma^2$

ประมาณช่วงความเชื่อมั่น (1-α) ของ (μ_1-μ_2) โดยใช้ T และใช้ค่าความแปร ปรวนรวมที่คำนวณจากตัวอย่าง (Sample Pooled Variance, Sp)

กรณี 2 ทาง

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad ...(4.64)$$

เมื่อ $Sp^2 = \dfrac{(n_1 -1) S_1^2 + (n_2 -1) S_2^2}{(n_1 + n_2 -2)}$ ....(4.65)

และ $\upsilon = n_1+n_2-2$ ....(4.66)

ตัวอย่างที่ 4.19 ในการขุดเจาะสำรวจก๊าซธรรมชาติ 2 หลุด เพื่อวิเคราะห์หาเปอร์เซ็นต์ของกำมะถัน หลุมแรกสุ่ม 12 ตัวอย่าง ได้ค่าเปอร์เซ็นต์กำมะถันเฉลี่ยได้ 3.11% และส่วนเบี่ยงเบนมาตรฐาน 0.771% หลุมที่สองสุ่ม 10 ตัวอย่าง ได้ค่าเปอร์เซ็นต์กำมะถันเฉลี่ยได้ 2.04% และส่วนเบี่ยงเบนมาตรฐาน 0.448 % จงหาช่วงความเชื่อมั่น 90% ของผลต่างของค่าเฉลี่ยเปอร์เซ็นต์กำมะถันในก๊าซธรรมชาติของทั้งสองหลุม ถ้าเปอร์เซ็นต์กำมะถันของแหล่งแร่ดังกล่าวมีการแจกแจงปกติ และมีความแปรปรวนเท่ากัน

วิธีทำ ไม่ทราบ σ_1^2 และ σ_2^2 แต่การสุ่มตัวอย่างจากทั้งสองหลุมซึ่งมาจากแหล่งเดียวกัน จึงเชื่อว่า $\sigma_1^2 = \sigma_2^2$ n_1 และ $n_2 < 30$

ประมาณช่วงความเชื่อมั่น 90% ของ $(\mu_1-\mu_2)$ โดยใช้ T แบบ 2 ทาง

$$(\bar{x}_1-\bar{x}_2) = 3.11-2.04 = 1.07\%$$

$$s_p^2 = \frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{(n_1+n_2-2)}$$

$$s_p^2 = \frac{11x771^2 + 9 \text{ x } 0.448^2}{(12+10-2)}$$

$$s_p^2 = 0.417$$

$$\upsilon = n_1+n_2-2 = 12+10-2$$

$$\upsilon = 20$$

ช่วงความเชื่อมั่น 90% ของ $\mu_1-\mu_2$ คือ

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} \cdot s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}}\, s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(3.11-2.04)\text{-}1.725\text{x}0.646\sqrt{\frac{1}{12}+\frac{1}{10}} \le \mu_1-\mu_2 < (3.11\text{-}2.04)+1.725\text{x}0646.\sqrt{\frac{1}{12}+\frac{1}{10}}$$

$$0.593 \le \mu_1-\mu_2 < 1.547$$

สรุป ผลต่างของค่าเฉลี่ยเปอร์เซ็นต์กำมะถันในก๊าซธรรมชาติของทั้งสองหลุมอยู่ระหว่าง 0.593% ถึง 1.547% ด้วยความเชื่อมั่น 90%

4.10.4 ถ้าไม่ทราบ σ_1^2, σ_2^2 แต่ n_1 และ $n_2 < 30$ และ $\sigma_1^2 \neq \sigma_2^2$

ประมาณช่วงความเชื่อมั่น $(1-\alpha)$ ของ $(\mu_1-\mu_2)$ โดยใช้ T

กรณี 2 ทาง

$$(\bar{x}1 - \bar{x}2) - t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 < (\bar{x}1 - \bar{x}2) + t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \quad ....(4.67)$$

เมื่อ $$\upsilon = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(\frac{S_1^2}{n_1})^2 (\frac{1}{n_1-1}) + (\frac{S_2^2}{n_2})^2 (\frac{1}{n_2-1})} \quad ....(4.68)$$

ตัวอย่างที่ 4.20 จากตัวอย่างที่ 4.19 ถ้าการขุดเจาะสำรวจก๊าซธรรมชาติของหลุมแรก และหลุมที่สอง พบว่าเป็นก๊าซธรรมชาติคนละแหล่ง และแต่ละหลุมมีการแจกแจงเปอร์เซ็นต์กำมะถันแบบปกติที่มีค่าความแปรปรวนไม่เท่ากันและไม่ทราบค่า จงหาช่วงความเชื่อมั่น 90% ของผลต่างของค่าเฉลี่ยเปอร์เซ็นต์กำมะถันในก๊าซธรรมชาติจากสองหลุมดังกล่าว

วิธีทำ ไม่ทราบค่า σ_1^2 และ σ_2^2 และ $\sigma_1^2 \neq \sigma_2^2$ และ n_1 และ $n_2 < 30$ ประมาณช่วงความเชื่อมั่นของ $(\mu_1-\mu_2)$ โดยใช้ T แบบ 2 ทาง

$(\bar{x}_1-\bar{x}_2)$ = 3.11-2.04 = 1.07%

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\left[\frac{(s_1^2/n_1)^2}{(n_1-1)}+\frac{(s_2^2/n_2)^2}{(n_2-1)}\right]}$$

$$= \frac{\left(0.771^2/12 + 0.448^2/10\right)^2}{\left[\frac{(0.771^2/12)^2}{(11)} + \frac{(0.448^2/10)^2}{(9)}\right]} = 18$$

ช่วงความเชื่อมั่น 90% ของ μ_1-μ_2 คือ

$$(\bar{x}1 - \bar{x}2) - t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 < (\bar{x}1 - \bar{x}2) + t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(3.11-2.04) - 1.734.\sqrt{\frac{0.771^2}{12} + \frac{0.448^2}{10}} \le \mu_1 - \mu_2 < (3.11\text{-}2.04) + 1.734.\sqrt{\frac{0.771^2}{12} + \frac{0.448^2}{10}}$$

$$0.6125 \le \mu_1 - \mu_2 < 1.5275$$

4.10.5 สรุปการหาช่วงความเชื่อมั่นของ μ หรือของ (μ_1-μ_2) กรณีไม่ทราบ σ^2

- **ถ้า n > 30**

ช่วงความเชื่อมั่น (1-α) = Point Estimate $\pm z_{\frac{\alpha}{2}} \hat{S}e$ ....(4.69)

เมื่อ $\hat{S}e$ = ค่าความผิดพลาดมาตราฐาน (Standard Error) ของ Point Estimate

$= \sqrt{\text{Var}(\bar{X} - \mu)} = \sqrt{\text{Var}(\bar{X})}$ ....(4.70)

- **ถ้า n < 30**

ช่วงความเชื่อมั่น (1-α) = Point Estimate $\pm t_{\frac{\alpha}{2}} \hat{S}e$ ....(4.71)

4.11 ช่วงความเชื่อมั่นของค่าเฉลี่ยของผลต่างของค่าที่สุ่มเป็นคู่ (Paired Observation)

การสุ่มตัวอย่างเป็นคู่ (Paired Observation) จากประชากรชุดเดียวกัน เช่น สุ่มวัดน้ำหนัก คน ๆ เดียวกัน ก่อนและหลังเข้าโครงลดน้ำหนัก หรืออาการคนใช้ก่อนและหลังใช้ยา ถือเป็นการสุ่มตัวอย่างที่

ไม่เป็นอิสระต่อกัน ซึ่งต่างจากการสุ่มหาช่วงความเชื่อมั่นของผลต่างของค่าเฉลี่ยของประชากร 2 ชุดในหัวข้อ 7.4

ประมาณช่วงความเชื่อมั่น (1-α) ของค่าเฉลี่ยของผลต่างของค่าที่สุ่มเป็นคู่โดยใช้ T
กรณี 2 ทาง

$$\bar{d}-t_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}} \le \mu_d < \bar{d}+t_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}} \qquad \text{..........................(4.72)}$$

ตัวอย่างที่ 4.21 ในการตรวจสอบสารอาหารชนิดหนึ่งในมะเขือเทศสด และมะเขือเทศแช่อิ่ม ได้ทำการสุ่มตัวอย่างมะเขือเทศสด ๆ ไปตรวจสอบสารอาหาร แล้วจึงนำไปทำมะเขือเทศแช่อิ่ม และตรวจสอบสารอาหารดังกล่าวอีกครั้งหนึ่ง สุ่มตัวอย่างทั้งหมด 10 ตัวอย่าง ได้ผลดังนี้

จงประมาณค่าเฉลี่ยของผลต่างที่แท้จริงของสารอาหารดังกล่าว ก่อนและหลังจากแช่อิ่ม ที่ระดับความเชื่อมั่น 98%

ตัวอย่างคู่ที่ (i)	ปริมาณสารอาหาร		
	มะเขือเทศสด	มะเขือเทศแช่อิ่ม	ผลต่าง (di)
1	0.066	0.085	0.019
2	0.079	0.088	0.009
3	0.069	0.091	0.022
4	0.076	0.096	0.020
5	0.071	0.093	0.022
6	0.089	0.095	0.006
7	0.071	0.079	0.008
8	0.073	0.078	0.005
9	0.069	0.065	-0.004
10	0.062	0.068	0.006
		$\bar{d}$	0.0117

วิธีทำ หาช่วงความเชื่อมั่น 98% ของ μ_d จากค่า $\bar{d}$

$$S_d = \sqrt{\frac{n\sum_{i=1}^{n} d_i^2 - (\sum_{i=1}^{n} d_i)^2}{n(n-1)}}$$

$$= \sqrt{\frac{10 \text{ x } 0.002003 - 0.117^2}{10 \text{ x } 9}}$$

$= \quad 0.0084$

$\alpha \quad = \quad 0.02$ และ $\upsilon \quad = \quad n-1 \; = 9$

ดังนั้น $t_{0.01,9} \quad = \quad 2.821$

ช่วงความเชื่อมั่น 98% ของ μ_d กรณี 2 ทางคือ

$$\bar{d}-t_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}} \le \mu_d < \bar{d}+t_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}}$$

$$0.0117-2.821 \text{ x } \frac{0.0084}{\sqrt{10}} \le \mu_d < 0.0117+2.821\frac{0.0084}{\sqrt{10}}$$

$$0.0042 \le \mu_d < 0.0192$$

สรุป ค่าเฉลี่ยของผลต่างที่แท้จริง (ประชากร) ของสารอาหารในมะเขือเทศสดและแช่อิ่ม จะอยู่ระหว่าง 0.0042 ถึง 0.0192 ด้วยความเชื่อมั่น 98%

4.12 ช่วงความเชื่อมั่นของสัดส่วน (Proportion , p)

สัดส่วน p คือค่าพารามิเตอร์ของการแจกแจงแบบทวินาม ซึ่งค่าประมาณของ p คือ $\hat{p} = \frac{x}{n}$

$\hat{P}$ คือตัวแปรสุ่มที่แสดงค่าสัดส่วนซึ่งค่าเฉลี่ยเท่ากับ p และความแปรปรวนเท่ากับ $\frac{pq}{n}$

- **ถ้า n > 30 ประมาณช่วงความเชื่อมั่นของ p โดยใช้ Z**

กรณี 2 ทาง

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \le p < \hat{p}+z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \qquad \text{..................................(4.73)}$$

ในการประมาณค่าจะไม่ทราบ p, q จึงใช้ $\hat{p}$, $\hat{q}$ แทน

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p < \hat{p}+z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad \text{..................................(4.74)}$$

ตัวอย่างที่ 4.22 ในการสุ่มตัวอย่างครอบครัวใน กทม. 500 ครอบครัว พบว่ามีเครื่องซักผ้าจำนวน 340 ครอบครัว จงประมาณค่าช่วงความเชื่อมั่น 95% ของสัดส่วนที่แท้จริงของครอบครัวใน กทม. ที่มีเครื่องซักผ้า (p)

วิธีทำ ตัวประมาณค่าแบบจุดของสัดส่วนที่แท้จริงของครอบครัวใน กทม. ที่มีเครื่องซักผ้าคือ

$$\hat{p} = \frac{x}{n} = \frac{340}{500} = 0.68$$

กรณี n > 30 ประมาณช่วงความเชื่อมั่นของ p โดยใช้ z

ที่ระดับความเชื่อมั่น 95%, $z_{.025} = 1.96$

ช่วงความเชื่อมั่น 95% ของ p กรณี 2 ทางคือ

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.68 - 1.96\sqrt{\frac{0.68 \times 0.32}{500}} \le p < 0.68 + 1.96\sqrt{\frac{0.68 \times 0.32}{500}}$$

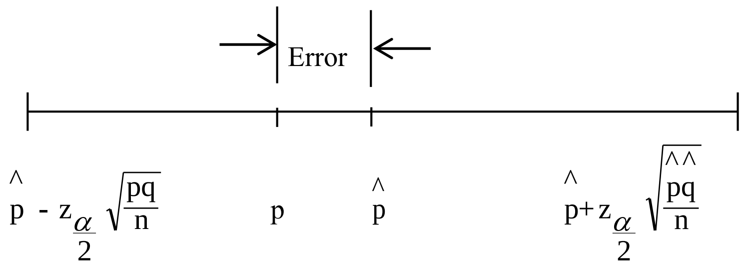
$$0.64 \le p < 0.72$$

สรุป สัดส่วนของประชากรที่มีเครื่องซักผ้าอยู่ในช่วง 0.64 ถึง 0.72 ด้วยความเชื่อมั่น 95%

- **การกำหนดขนาดของตัวอย่างที่เหมาะสม (Appropriate Sample Size)**

ความผิดพลาดในการใช้ $\hat{p}$ ประมาณค่า p ของประชากร (รูปที่ 4.12) เท่ากับ

$$\text{Error} = |\hat{p} - p| = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad \text{.........................(4.75)}$$



รูปที่ 4.12 ความผิดพลาดในการประมาณค่า p

ในทำนองเดียวกับสมการ (4.61) สามารถสรุปได้ว่า ถ้า $\hat{p}$ เป็นตัวประมาณค่าของ p ค่าความผิดพลาดจะไม่เกินค่า e ที่ระดับความเชื่อมั่นได้ $(1-\alpha)$ 100%

$$n \ge \frac{z^2_{\alpha/2}\,\hat{p}\hat{q}}{e^2} \qquad \text{.........................(4.76)}$$

สมการที่ 4.76 ใช้ค่า $\hat{p}$ เพื่อหาขนาดของตัวอย่าง n โดยหา $\hat{p}$ จากตัวอย่าง ในกรณีที่ไม่ทราบค่า p ของประชากร โดยอาจสุ่มตัวอย่างขนาดมากกว่า 30 ตัวอย่างเพื่อหาค่า $\hat{p}$ แต่ถ้าทราบค่า p ของประชากรควรใช้ค่า p ของประชากร ค่าของ n จะเป็นจำนวนเต็มและปัดเศษขึ้นเสมอ

ตัวอย่างที่ 4.23 ถ้าสุ่มตัวอย่างมา 100 ชิ้น พบว่ามีผลิตภัณฑ์บกพร่อง 8 ชิ้น จะต้องใช้ขนาดตัวอย่าง (n) เท่าใดในการประมาณค่าสัดส่วนของผลิตภัณฑ์ที่มีข้อบกพร่อง (p) ที่ระดับความเชื่อมั่น 98% ถ้าต้องการความผิดพลาดไม่เกิน 0.01

วิธีทำ

$$\hat{p} = \frac{8}{100}, \qquad \hat{q} = 0.92$$

$$z_{0.01} = 2.325$$

$$n = \frac{z^2_{\alpha/2}\,\hat{p}\hat{q}}{e^2}$$

$$= \frac{(2.325)^2 \text{ x } 0.08 \text{ x } 0.92}{0.01^2}$$

$$= 39.78$$

ขนาดของตัวอย่าง (n) = 40 ตัวอย่าง

4.13 ช่วงความเชื่อมั่นของผลต่างของสัดส่วน (p_1-p_2)

- **ถ้า $n > 30$ ประมาณช่วงความเชื่อมั่นของ (p_1-p_2) โดยใช้ Z**

  กรณี 2 ทาง

$$(\hat{p}_1-\hat{p}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} \le p_1-p_2 < (\hat{p}_1-\hat{p}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} \quad ....(4.77)$$

ตัวอย่างที่ 4.24 ก่อนที่บริษัทจัดโปรแกรมโฆษณาสินค้า ได้ทำการสุ่มตัวอย่างลูกค้า 100 คน พบว่า 16 คน ชอบเครื่องดื่มชนิดหนึ่ง ภายหลังจากที่บริษัทจัดโปรแกรมโฆษณาสินค้าได้สุ่มตัวอย่างลูกค้า 200 คน ปรากฏว่า 80 คนชอบเครื่องดื่มชนิดนั้น อยากทราบว่าโฆษณามีผลทำให้คนนิยมดื่มเครื่องดื่มชนิดนั้นหรือไม่ โดยพิจารณาจากช่วงความเชื่อมั่น 95% ของความแตกต่างระหว่างสัดส่วนของผู้ที่ชอบเครื่องดื่มชนิดนี้ก่อน และหลังโฆษณา

วิธีทำ

$$\hat{p}_1 = 0.16 \qquad \hat{q}_1 = 0.84$$

$$\hat{p}_2 = 0.40 \qquad \hat{q}_2 = 0.60$$

$$(1-\alpha) = 0.95 \qquad z_{0.025} = 1.96$$

ช่วงความเชื่อมั่น 95% ของ (p_2-p_1) คือ

$$(\hat{p}_2-\hat{p}_1) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}+\frac{\hat{p}_1\hat{q}_1}{n_1}} \le p_2-p_1 < (\hat{p}_2-\hat{p}_1)+z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}+\frac{\hat{p}_1\hat{q}_1}{n_1}}$$

$$(0.40-0.16)-1.96\sqrt{\frac{0.4 \text{ x } 0.6}{200}+\frac{0.16 \text{ x } 84}{100}} \le p_2-p_1 < (0.40-0.16)+1.96\sqrt{\frac{0.4 \text{ x } 0.6}{200}+\frac{0.16 \text{ x } 84}{100}}$$

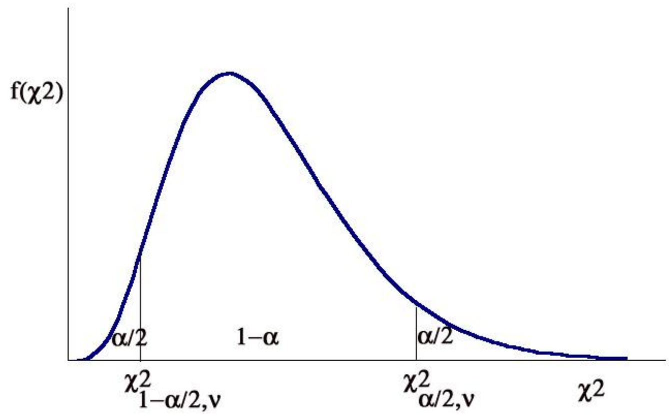
$$0.14114 \le (p_2-p_1) < 0.33886$$

สรุป ค่าประมาณของ p_2-p_1 มากกว่าศูนย์ หรือ p_2>p_1 แสดงว่าการโฆษณามีผลทำให้สัดส่วนของผู้ที่ชอบเครื่องดื่มชนิดนี้สูงขึ้น ด้วยระดับความเชื่อมั่น 95%

4.14 ช่วงความเชื่อมั่นของ σ^2

- **ประมาณช่วงความเชื่อมั่นของ σ^2 โดยใช้ χ^2**

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}$$ มีการแจกแจงดังรูปที่ 4.13



รูปที่ 4.13 การแจกแจงของ χ^2

จากรูปการแจกแจงของ χ^2 จะสามารถหาช่วงความเชื่อมั่น (1-α) ของ σ^2 ได้ดังนี้

$$P(\chi^2_{1-\frac{\alpha}{2},\upsilon} \le \frac{(n-1)\,S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2},\upsilon}) = 1-\alpha \quad \text{.......................(4.78)}$$

$$P(\frac{(n-1)\,S^2}{\chi^2_{1-\frac{\alpha}{2},\upsilon}} > \sigma^2 \ge \frac{(n-1)\,S^2}{\chi^2_{\frac{\alpha}{2},\upsilon}}) = 1-\alpha \quad \text{.......................(4.79)}$$

$$P\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},\upsilon}} \le \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},\upsilon}}\right] = 1-\alpha \qquad \text{.......................(4.80)}$$

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},\upsilon}} \le \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},\upsilon}} \qquad \text{.......................(4.81)}$$

ตัวอย่างที่ 4.25 สุ่มตัวอย่างสินค้าจำนวน 10 หีบห่อที่ผลิตได้ในวันเดียวพบว่าน้ำหนักของสินค้าที่บรรจุอยู่ในหีบห่อมีค่าเท่ากับ 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2 และ 46.0 ออนซ์ จงหาช่วงความเชื่อมั่น 95% ของค่าความแปรปรวนของน้ำหนักสินค้าที่ผลิตได้ในวันนั้นทั้งหมด ถ้าน้ำหนักมีการแจกแจงปกติ

วิธีทำ

$$\upsilon = 10-1 = 9$$

$$S^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

$$= \frac{10 \times 21{,}273.12 - (461.2)^2}{10 \times 9}$$

$$= 0.286$$

ช่วงความเชื่อมั่น 95% ของ σ^2 คือ

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} \le \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}$$

$$\frac{(10-1)\times 0.286}{\chi^2_{0.025,\,9}} \le \sigma^2 < \frac{(10-1)\times 0.286}{\chi^2_{0.975,\,9}}$$

$$\frac{9 \times 0.286}{19.023} \le \sigma^2 < \frac{9 \times 0.286}{2.700}$$

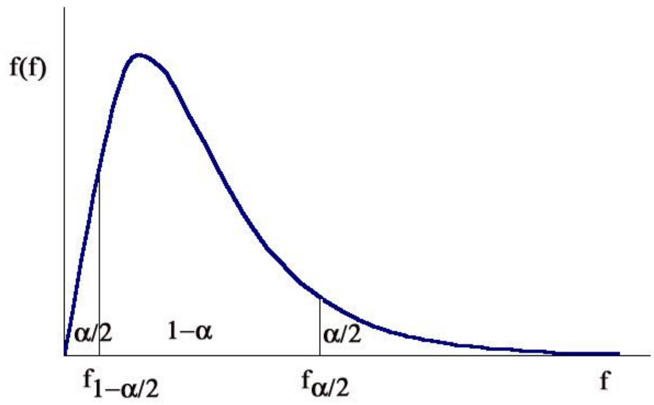
$$0.135 \le \sigma^2 < 0.953$$

สรุป ค่าความแปรปรวน (σ^2) ของน้ำหนักสินค้าที่ผลิตได้ในวันนั้นทั้งหมดมีค่าอยู่ระหว่าง 0.135 ออนซ์ถึง 0.953 ออนซ์ ด้วยความเชื่อมั่น 95%

4.15 ช่วงความเชื่อมั่นของอัตราส่วนของ σ^2_1/σ^2_2

➢ **ประมาณช่วงความเชื่อมั่นของ σ^2_1/σ^2_2 โดยใช้ F**

$$F = \frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}$$ มีการแจกแจงดังรูปที่ 4.14



รูปที่ 4.14 การแจกแจงของ F

จากรูปการแจกแจงของ F จะสามารถหาช่วงความเชื่อมั่น (1-α) ของ $\frac{\sigma_1^2}{\sigma_2^2}$ ให้ดังนี้

$$P\left(f_{1-\frac{\alpha}{2}} \le \frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} < f_{\frac{\alpha}{2}}\right) = 1-\alpha \qquad \text{.....................(4.82)}$$

$$P\left(f_{1-\frac{\alpha}{2}}\ \frac{S_2^2}{S_1^2} \le \frac{\sigma_2^2}{\sigma_1^2} < f_{\frac{\alpha}{2}}\cdot\frac{S_2^2}{S_1^2}\right) = 1-\alpha$$

$$P\left(\frac{S_1^2}{S_2^2}\ \frac{1}{f_{1-\frac{\alpha}{2}}} > \frac{\sigma_1^2}{\sigma_2^2} \ge \frac{S_1^2}{S_2^2\ f_{\frac{\alpha}{2}}}\right) = 1-\alpha$$

$$P\left(\frac{s_1^2}{s_2^2}\frac{1}{f_{\frac{\alpha}{2},(\upsilon_1,\upsilon_2)}} \le \frac{\sigma_1^2}{\sigma_2^2} <. \frac{s_1^2}{s_2^2}\frac{1}{f_{1-\frac{\alpha}{2},(\upsilon_1,\upsilon_2)}}\right) = 1-\alpha$$

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\frac{\alpha}{2},(\upsilon_1,\upsilon_2)}} \le \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}.\, f_{\frac{\alpha}{2},(\upsilon_2,\upsilon_1)} \qquad ....(4.83)$$

ตัวอย่างที่ 4.26 ต้องการเปรียบเทียบกำลังส่องสว่างของหลอดไฟฟ้าชนิด 2 ชนิด โดยการทดสอบหลอดไฟฟ้าชนิดที่ 1 16 ครั้งได้ $s_1 = 2.7$ แรงเทียน หลอดไฟฟ้าชนิดที่ 2 21 ครั้ง ได้ $s_2 = 4.2$ แรงเทียน จงหาช่วงความเชื่อมั่น 98% สำหรับ σ^2_2/σ^2_1

วิธีทำ

$$\frac{s_2^2}{s_1^2}\frac{1}{f_{\frac{\alpha}{2},(\upsilon_2,\upsilon_1)}} \le \frac{\sigma_2^2}{\sigma_1^2} < \frac{s_2^2}{s_1^2}.\, f_{\frac{\alpha}{2},(\upsilon_1,\upsilon_2)}$$

$$\left(\frac{4.2}{2.7}\right)^2 \frac{1}{f_{0.01,\,(20,\,15)}} \le \frac{\sigma_2^2}{\sigma_1^2} < \left(\frac{4.2}{2.7}\right) f_{0.01,\,(15,\,20)}$$

$$\left(\frac{4.2}{2.7}\right)^2 \text{x } \frac{1}{3.37} \le \frac{\sigma_2^2}{\sigma_1^2} < \left(\frac{4.2}{2.7}\right) \text{x } 3.09$$

$$0.7181 \le \frac{\sigma_2^2}{\sigma_1^2} < 7.477$$

อัตราส่วน σ^2_2/σ^2_1 อยู่ระหว่าง 0.7181 และ 7.477 ด้วยความเชื่อมั่น 98%

4.16 สรุปการหาช่วงความเชื่อมั่น

สูตรการหาช่วงความเชื่อมั่นที่ $(1-\alpha)$ ของพารามิเตอร์ของประชากรจากค่าสถิติที่คำนวณจากตัวอย่าง ทั้งกรณีคำนวณช่วงความเชื่อมั่นแบบ 2 ทาง ทางเดียวด้านล่าง และทางเดียวด้านบน สรุปไว้ในตารางที่ 4.3

ตารางที่ 4.3 สรุปสูตรของช่วงความเชื่อมั่นที่ (1-α)

[กรณีประชากร 1 ชุด]			
พารามิเตอร์	เกณฑ์การพิจารณาเลือกค่าสถิติ 1	เกณฑ์การพิจารณาเลือกค่าสถิติ 2	ช่วงความเชื่อมั่นที่ (1-α)
μ	ทราบ σ^2	-	➢ $\bar{x}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu < \bar{x}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ ➢ $\bar{x}-z_{\alpha}\frac{\sigma}{\sqrt{n}} \le \mu < \infty$ ➢ $-\infty \le \mu < \bar{x}+z_{\alpha}\frac{\sigma}{\sqrt{n}}$
	ไม่ทราบ σ^2	n > 30	➢ $\bar{x}-z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \le \mu < \bar{x}+z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$
	ไม่ทราบ σ^2	n < 30	➢ $\bar{x}-t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \le \mu < \bar{x}+t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$ (υ = n-1)
p	-	n > 30	➢ $\hat{p}-z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p < \hat{p}+z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$
μ_d	ไม่ทราบ σ_d^2	n < 30	➢ $\bar{d}-t_{\frac{\alpha}{2}}\frac{Sd}{\sqrt{n}} \le \mu d < \bar{d}+t_{\frac{\alpha}{2}}\frac{Sd}{\sqrt{n}}$ (υ = n-1)
σ^2	-	-	➢ $\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$

[กรณีประชากร 2 ชุด]			
พารามิเตอร์	เกณฑ์การพิจารณาเลือกค่าสถิติ		ช่วงความเชื่อมั่นที่ (1-α)
	1	2	
μ_1-μ_2	ทราบ σ_1^2, σ_2^2	-	➢ $(\bar{x}_1-\bar{x}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 <$ $(\bar{x}_1-\bar{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
	ไม่ทราบ σ_1^2, σ_2^2	$n_1, n_2 > 30$	➢ $(\bar{x}_1-\bar{x}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \le \mu_1-\mu_2 <$ $(\bar{x}_1-\bar{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$
	ไม่ทราบ σ_1^2, σ_2^2 แต่ $\sigma_1^2 = \sigma_2^2$	$n_1, n_2 < 30$	➢ $(\bar{x}_1-\bar{x}_2) - t_{\frac{\alpha}{2}}\, Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu <$ $(\bar{x}_1-\bar{x}_2) + t_{\frac{\alpha}{2}}\, Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$; $\nu = n_1+n_2-2$
	ไม่ทราบ σ_1^2, σ_2^2 แต่ $\sigma_1^2 \neq \sigma_2^2$	$n_1, n_2 < 30$	➢ $(\bar{x}_1-\bar{x}_2) - t_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \le \mu_1-\mu_2 <$ $(\bar{x}_1-\bar{x}_2) + t_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$ $\nu = \frac{\left[\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right]^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}$
p_1- p_2	-	-	➢ $(\hat{p}_1-\hat{p}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} \le p_1-p_2 <$ $(\hat{p}_1-\hat{p}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$
$\frac{\sigma_1^2}{\sigma_2^2}$	-	-	$\frac{s_1^2}{s_2^2} f_{\frac{\alpha}{2},(\upsilon_1,\upsilon_2)} \le \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\frac{\alpha}{2},(\upsilon_2,\upsilon_1)}$ $\nu_1 = n_1-1$, $\nu_2 = n_2-1$

4.17 แบบฝึกหัด

1. ในการสุ่มวัดอุณหภูมิอากาศเฉลี่ยประจำวันของภูมิอากาศที่อำเภอแห่งหนึ่ง เป็นเวลา 15 วัน ในช่วงฤดูร้อน ถ้าสมมติว่าประชากรของอุณหภูมิอากาศเฉลี่ยมีการแจกแจงแบบปกติโดยมีค่าเฉลี่ย 35°C และส่วนเบี่ยงเบนมาตรฐาน 3°C จงหาค่าความน่าจะเป็นที่
   (ก) ค่าเฉลี่ยจากตัวอย่างจะมีค่ามากกว่า 40°C
   (ข) ค่าส่วนเบี่ยงเบนมาตรฐานมีค่ามากกว่า 5°C

2. มหาวิทยาลัยแห่งหนึ่งมีนิสิต 2,500 คน ถ้าความสูงของนิสิตในมหาวิทยาลัยมีการแจกแจงแบบปกติโดยมีค่าเฉลี่ย 163 ซม. และส่วนเบี่ยงเบนมาตราฐาน 5 ซม. ถ้าทำการสุ่มตัวอย่าง 60 ชุด ๆ ละ 30 คน จะมีตัวอย่างกี่ชุดที่มีความสูงเฉลี่ย
   (ก) ระหว่าง 160-170 ซม.
   (ข) น้อยกว่า 155 ซม.

3. ให้ $(x_1, x_2, \ldots\ldots, x_{10})$ และ $(y_1, y_2, \ldots\ldots, y_8)$ คือตัวอย่างที่สุ่มจากประชากรปกติ 2 ชุด ซึ่งเป็นอิสระต่อกัน และมีความแปรปรวนเท่ากับ
   เมื่อ X คือตัวแปรสุ่มแสดงจำนวนชิ้นงานที่ผลิตได้จากเครื่องจักร A ในเวลา 1 ชั่วโมง
   และ Y คือตัวแปรสุ่มแสดงจำนวนชิ้นงานประเภทเดียวกันที่ผลิตได้จากเครื่องจักร B ในเวลา 1 ชั่วโมง
   ถ้าผลการสุ่มตัวอย่างพบว่า

   $$\bar{x} = 10.5 \quad \text{และ} \quad \sum_{i=1}^{10} x_i^2 = 1{,}165$$

   $$\bar{y} = 9 \quad \text{และ} \quad \sum_{i=1}^{8} y_i^2 = 1{,}020$$

   จงหาความน่าจะเป็นที่โดยเฉลี่ยเครื่องจักร A จะผลิตชิ้นงานได้มากกว่าเครื่องจักร B

4. ถ้าสุ่มตัวอย่างจากประชากรชุดหนึ่งโดยใช้ขนาดตัวอย่าง 25 จะได้ค่าความคาดเคลื่อนมาตราฐาน (Standard Error หรือ Standard Deviation of Mean) เป็น 3 จงคำนวณหาว่าจะต้องใช้ขนาดตัวอย่างเท่าไร จึงจะทำให้ค่าความคาดเคลื่อนมาตราฐานลดลงเหลือ 1.5

5. นักวิจัยคนหนึ่งต้องการจะประมาณค่าเฉลี่ยของประชากรชุดหนึ่งโดยการสุ่มตัวอย่าง อยากทราบว่าเขาจะต้องสุ่มตัวอย่างขนาดเท่าใด จึงจะทำให้ผลต่างระหว่างค่าเฉลี่ยของตัวอย่างและค่าเฉลี่ยของประชากรมีค่าไม่เกิน 25% ของค่าส่วนเบี่ยงเบนมาตราฐานของประชากรด้วยความน่าจะเป็นเท่ากับ 0.95

6. ในการตรวจสอบผลิตภัณฑ์ชนิดหนึ่งมีการแจกแจงแบบปกติซึ่งมีผลิตภัณฑ์บกพร่อง 3% จงหาความน่าจะเป็นที่ในการสุ่มตรวจสอบผลิตภัณฑ์ชนิดนี้จำนวน 500 ชิ้น จะพบว่ามีผลิตภัณฑ์บกพร่อง

   (ก) 4% หรือมากกว่า

   (ข) 2% หรือน้อยกว่า

7. ถ้า U คือตัวแปรสุ่มที่มีการแจกแจงปกติ

   V คือตัวแปรสุ่มที่มีการแจกแจงแบบไคสแควร์ด้วยองศาเสรี 9

   W คือตัวแปรสุ่มที่มีการแจกแจงแบบไคสแควร์ด้วยองศาเสรี 11

   และให้ U, V และ W เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน จงระบุว่าตัวแปรสุ่มต่อไปนี้มีการแจกแจงแบบใด พร้อมคำนวณพารามิเตอร์และองศาเสรี

   (ก) $X = \dfrac{3U}{\sqrt{V}}$

   (ข) $Y = \dfrac{11V}{9W}$

8. จงหา $z_\alpha, t_\alpha, \chi^2_\alpha, \chi^2_{1-\alpha}, f_\alpha$ และ $f_{1-\alpha}$ ถ้า

   (ก) $P(|Z| > z_\alpha) = 0.1164$

   (ข) $P(|T| > t_\alpha) = 0.05$ เมื่อ $n = 20$

   (ค) $P(\chi^2_{1-\alpha} < \chi^2 < \chi^2_\alpha) = 0.9$ เมื่อ $\upsilon = 17$

   (ง) $P(f_{1-\alpha} < F < f_\alpha) = 0.9$ เมื่อ $\upsilon_1 = 10$; $\upsilon_2 = 15$

9. ในการสุ่มตรวจวัดการให้พลังงานความร้อน (ล้านแคลอรี่ต่อตัน) ของถ่านหินจากเหมืองถ่านหิน 2 เหมืองได้ผลดังนี้

เหมืองที่	พลังงานความร้อน (ล้านแคลอรี่ต่อตัน)					
1	8,260	8,130	8,350	8,070	8,340	
2	7,950	7,890	7,900	8,140	7,920	7,840

อยากทราบว่าจะสามารถสรุปว่าค่าความแปรปรวนของถ่านหินทั้ง 2 เหมืองมีค่าเท่ากันได้หรือไม่

10. ประชากร 2 ชุด มีค่าความแปรปรวน σ_1^2 และ σ_2^2 และทราบว่า $\sigma_1^2 = \sigma_2^2$ ทำการสุ่มตัวอย่างจากประชากรชุดที่ 1 โดยกำหนด $n_1 = 10$ ได้ค่าความแปรปรวน S_1^2 และสุ่มตัวอย่างจากประชากรชุดที่ 2 โดยกำหนด $n_2 = 25$ ได้ค่าความแปรปรวน S_2^2

ถ้า $P\left(\frac{S_1^2}{S_2^2} < b\right) = 0.95$ จงหาค่า b

11. วิศวกรโยธาผู้หนึ่งต้องการวัดความสามารถรับแรงอัดของคอนกรีตสองชนิด โดยทำการสุ่มตัวอย่างคอนกรีตชนิดละ 10 ตัวอย่างมาทำการทดสอบ (หน่วยเป็น ksc) ได้ข้อมูลดังตาราง

ความสามารถรับแรงอัดของคอนกรีต (ksc)	
ชนิดที่ 1	ชนิดที่ 2
229	218
230	219
303	212
224	216
230	210
232	220
235	234
247	226
216	238
220	213

กำหนดให้ตัวอย่างทั้งหมดสุ่มจากประชากรที่มีการแจกแจงปกติสองชุดที่ค่าความแปรปรวนเท่ากัน จงหาค่าของ

(ก) ค่าประมาณแบบจุดของค่าเฉลี่ยและความแปรปรวนของความสามารถรับแรงอัดของคอนกรีตทั้งสองชนิด

(ข) ช่วงความเชื่อมั่น 95% ของผลต่างของค่าเฉลี่ยของความสามารถรับแรงอัดของคอนกรีตทั้งสองชนิด (μ_1-μ_2)

(ค) ช่วงความเชื่อมั่น 95% ของลิมิตล่างของค่าเฉลี่ยของความสามารถรับแรงอัดของคอนกรีตทั้งสองชนิด

(ง) ช่วงความเชื่อมั่น 99% ของลิมิตบนของค่าเฉลี่ยของความสามารถรับแรงอัดของคอนกรีตทั้งสองชนิด

12. กำหนดให้สัญญาณความถี่ของการสื่อสารส่งจากเมือง A ไปยังเมือง B มีการแจกแจงปกติมีค่าเฉลี่ย μ และความแปรปรวนเท่ากับ 4 ถ้าส่งสัญญาณชนิดนี้จากเมือง A ไปยังเมือง B 9 ครั้ง ผลการรับความถี่ของสัญญาณที่เมือง B ได้ค่าดังนี้ 5, 8.5, 12, 15, 7, 9, 7.5, 6.5 และ 10.5

(ก) จงสร้างช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยของค่าความถี่ของสัญญาณสื่อสารนี้

(ข) ถ้าต้องการให้ความผิดพลาดจากค่าเฉลี่ยจากตัวอย่างของค่าสัญญาณอยู่ในช่วงความกว้าง 0.1 ด้วยความเชื่อมั่น 99% อยากทราบว่า จะต้องสุ่มตัวอย่างขนาดเท่าใด

13. จากการศึกษาปริมาณความเข้มข้นของก๊าซคาร์บอนมอนนอกไซด์บริเวณศูนย์กลางของเมือง ๆ หนึ่ง โดยทำการเก็บอากาศตัวอย่างใส่ในถุงพิเศษ แล้วนำไปหาความเข้มข้นของก๊าซคาร์บอนมอนนอกไซด์ โดยใช้เครื่องสเปคโตรโฟโตมิเตอร์ (Spectrophotometer) ปริมาณความเข้มข้นที่วัดได้มีหน่วยเป็น ppm (part per million) ผลการเก็บตัวอย่างในช่วง 1 ปี จำนวน 20 ถุงตัวอย่าง วัดความเข้มข้นของก๊าซได้ผลดังนี้

104.1	101.1	102.2	100.4	98.6	88.2	78.8	83.0	84.7	94.8
105.1	106.2	111.2	108.3	105.2	103.2	99.0	98.8	102.2	98.4

จงสร้างช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยจริงของความเข้มข้นของก๊าซคาร์บอนมอนนอกไซด์ในบริเวณศูนย์กลางของเมืองดังกล่าว

14. ในการทดลองผลการใช้ยาลดน้ำหนักชนิดหนึ่ง ได้ทำการสุ่มตัวอย่างคนจำนวน 10 คน บันทึกน้ำหนักก่อนใช้ยาไว้ และทำการบันทึกอีกครั้งหนึ่งแล้วจากใช้ยาเป็นเวลา 3 เดือน ได้ผลลัพธ์ดังตาราง (หน่วยเป็นกิโลกรัม)

คน	ก่อนใช้ยา	หลังใช้ยา
1	52	48
2	51	50
3	55	52
4	54	60
5	49	50
6	57	53
7	56	58
8	60	50
9	55	48
10	52	55

จงสร้างช่วงความเชื่อมั่น 99% ของค่าเฉลี่ยของผลต่างของน้ำหนักก่อนและหลังใช้ยา

15. ในการผลิตสินค้าชนิดหนึ่ง โรงงานแห่งหนึ่งใช้สายการผลิตสองสาย คือ สาย A และสาย B เพื่อทดสอบค่าเฉลี่ยของจำนวนสินค้าที่ผลิตได้ใน 1 ชั่วโมงของสายการผลิตทั้งสองว่ามีความแตกต่างกันหรือไม่ วิศวกรผู้หนึ่งได้ทำการสุ่มนับจำนวนสินค้าที่ผลิตได้จากแต่ละสายการผลิตได้ข้อมูลจำนวนชิ้นสินค้าที่ผลิตได้ต่อชั่วโมงดังตาราง

สายการผลิต A	86	87	56	93	84	93	75	79
สายการผลิต B	80	79	58	91	77	82	74	66

สมมติว่า จำนวนที่ผลิตได้จากสายการผลิตทั้งสองมีการแจกแจงแบบปกติด้วยค่าความแปรปรวนของประชากรเท่ากัน จงสร้างช่วงความเชื่อมั่น 95% ของผลต่างของค่าเฉลี่ยของจำนวนสินค้าที่ผลิตได้จากสายการผลิตทั้งสองในหนึ่งชั่วโมง

16. คะแนนการสอบภาษาอังกฤษของนักศึกษากลุ่มหนึ่งจำนวน 600 คน ได้ผลสรุปดังตารางต่อไปนี้ (ตัวเลขในตารางแสดงจำนวนนักเรียน)

	ผลการสอบครั้งที่ 1			
	A	B	C	รวม

ผลการสอบ ครั้งที่ 2	A	140	60	20	220
	B	20	200	20	240
	C	30	60	50	140
รวม		190	320	90	600

ก. กำหนดให้ p_1 = สัดส่วนของนักเรียนที่ผลการสอบครั้งที่ 1 สูงกว่าครั้งที่ 2

p_2 = สัดส่วนของนักเรียนที่ผลการสอบครั้งที่ 2 สูงกว่าครั้งที่ 1

p_3 = สัดส่วนของนักเรียนที่ผลการสอบทั้งสองครั้งเท่ากัน

จงหาค่าประมาณแบบจุดของ p_1 , p_2 และ p_3

ข. จงสร้างความเชื่อมั่น 95% ของค่าสัดส่วนของนักเรียนที่มีการพัฒนาในผลการเรียน

ค. จงสร้างช่วงความเชื่อมั่น 95% ของผลต่างระหว่างค่าสัดส่วนของนักเรียนที่มีการพัฒนาใน ผลการเรียน และค่าสัดส่วนของนักเรียกที่มีผลการเรียนแย่ลง

17. วิศวกรผู้หนึ่งได้ทำการศึกษาเครื่องมือที่ใช้ในการวัดผลจากกระบวนเคมี 2 แบบ และมีข้อสงสัยว่าเครื่องมือเดิม (เครื่องมือชนิดที่ 1) มีค่าความแปรปรวนสูงกว่าเครื่องมือชนิดใหม่ (เครื่องมือชนิดที่ 2) จึงทำการสุ่มตัวอย่างเก็บรวบรวมข้อมูลโดยใช้เครื่องมือชนิดที่ 1 (ขนาด n_1) และเครื่องมือชนิดที่ 2 (ขนาด n_2) ได้ผลดังนี้

n_1 = 12 $\quad S_1^2$ = 14.5

n_2 = 10 $\quad S_2^2$ = 10.8

ก. จงสร้างช่วงความเชื่อมั่น 98% ของ σ_1^2 / σ_2^2

ข. จงสร้างช่วงความเชื่อมั่น 95% ของลิมิตบนของ σ_2 / σ_1

18. เครื่องจักร 2 เครื่องใช้ในการบรรจุน้ำดื่มลงในขวดพลาสติกที่มีความจุ 16 ออนซ์ ถ้ากระบวนการบรรจุน้ำดื่มนี้มีการแจกแจงปกติ ด้วยค่าส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 0.015 และ 0.018 ออนซ์ ตามลำดับ วิศวกรควบคุมคุณภาพการผลิตคนหนึ่งสงสัยว่าเครื่องจักรทั้งสองจะบรรจุน้ำดื่มในปริมาณที่เท่ากันหรือไม่ จึงทำการสุ่มตัวอย่างและรวบรวมข้อมูลปริมาณน้ำดื่มที่บรรจุขวดโดยเครื่องจักรทั้งสองได้ค่าดังตาราง

เครื่องจักร 1		**เครื่องจักร 2**	
16.03	16.01	16.02	15.99

16.04	15.96	15.97	16.03
16.05	15.98	15.96	16.04
16.05	16.02	16.01	
16.02	15.99	16.02	

จงสร้างช่วงความเชื่อมั่น 90% ของ $\mu_1-\mu_2$

19. บริษัทแทกซี่กำลังอยู่ระหว่างการตัดสินใจว่าซื้อยางรถยนต์ยี่ห้อ A หรือ B สำหรับแทกซี่ของบริษัท จึงได้ทำการทดสอบการใช้งานของยางทั้ง 2 ยี่ห้อ จำนวนอย่างละ 12 ชุด โดยการนำยางมาทดลองใช้จนกระทั่งยางหมดอายุซึ่งผลการทดสอบมีดังนี้

	ยาง A	ยาง B
$\bar{x}$ (กม.)	36,300	38,100
S (กม.)	5,000	6,100

จงคำนวณหาช่วงความเชื่อมั่น 95% ของ $\mu_1-\mu_2$ ถ้าสมมติว่าประชากรมีการแจกแจงปกติ

20. บริษัทผู้ผลิตแบตเตอรี่รถยนต์โฆษณาว่าแบตเตอรี่ของบริษัทมีอายุการใช้งานเฉลี่ย 3 ปี มีค่าความแปรปรวน 1 ปี ถ้านำแบตเตอรี่ของบริษัทดังกล่าวจำนวน 5 ลูกมาทดสอบอายุการใช้งานและได้ผลการทดสอบดังนี้ 1.9, 2.4, 3.0, 3.5, และ 4.2 ปี

จงสร้างช่วงความเชื่อมั่น 95% ของ σ^2 และสรุปว่า σ^2 เท่ากับ 1 จริงหรือไม่ ถ้าสมมติว่าอายุการใช้งานของแบตเตอรี่ดังกล่าวมีการแจกแจงแบบปกติ

บทที่ 5
การทดสอบสมมติฐาน (Hypothesis Testing)

5.1 คำนำ

เนื้อหาในบทที่ 4 เกี่ยวข้องกับการสุ่มตัวอย่าง การคำนวณค่าสถิติ การวิเคราะห์การแจกแจงความน่าจะเป็นของค่าสถิติที่เรียกว่าการแจกแจงตัวอย่างสุ่ม (Sampling Distribution) โดยมีข้อกำหนดต่าง ๆ กันคือ ในบทที่ 4 ได้กล่าวถึงการสุ่มตัวอย่างจากประชากรที่ไม่ทราบค่าพารามิเตอร์ แต่ต้องการคำนวณหาช่วงความเชื่อมั่น (Confidence Interval) ของค่าพารามิเตอร์ตัวนั้น ส่วนในบทนี้เป็นการตั้งสมมติฐานเกี่ยวกับค่าพารามิเตอร์ของประชากร แล้วทำการทดสอบว่าพารามิเตอร์มีค่าตามที่ตั้งสมมติฐานไว้หรือไม่

เช่น (1) อายุการใช้งานเฉลี่ยของยางรถยนต์ $>$ 30,000 กม.ตามที่บริษัทกำหนดไว้หรือไม่ [Ho : $\mu > 30{,}000$]

(2) การเพิ่มอัตราการใช้ปุ๋ยทำให้ผลผลิตข้าวต่อไร่เพิ่มหรือไม่ [H_o : $\mu_1 > \mu_2$]

(3) ร้อยละ 90 ของคนไข้ ใช้ยาชนิดนี้แล้วมีอาการดีขึ้นหรือไม่ [H_o : $p > 0.90$]

5.2 ขั้นตอนการทดสอบสมมติฐาน

การทดสอบสมมติฐานมีขั้นตอนดังต่อไปนี้

(1) การตั้งสมมติฐานหลัก (Null Hypothesis) หรือ H_o และสมมติฐานรอง (Alternative Hypothesis) หรือ H_1

(2) กำหนดระดับนัยสำคัญในการทดสอบ (Significant level, α) หรือค่าความน่าจะเป็นของการปฏิเสธสมมติฐานที่เป็นจริง (Type I Error) ซึ่งจะได้กล่าวถึงรายละเอียดในหัวข้อ 5.4 ปกติจะกำหนด α เท่ากับ 0.01 หรือ 0.05 หรือ 0.1

(3) สุ่มตัวอย่างและคำนวณค่าสถิติ (z, t, χ^2 , f)

(4) คำนวณหาช่วงวิกฤต (Critical Region, CR) และช่วงการยอมรับ (Acceptance Region , AC)

(5) ทดสอบสมมติฐาน

- ถ้าค่าสถิติอยู่ในช่วงวิกฤต ....ปฏิเสธ (Reject) H_o
- ถ้าค่าสถิติอยู่นอกช่วงวิกฤต ....ยอมรับ (Accept) H_o

5.3 วิธีการตั้งสมมติฐาน

(1) H_o (สมมติฐานหลัก, Null Hypothesis) หรือ

เช่น $H_o : \mu \geq 30{,}000$ กม.

$H_o : \mu_1 = \mu_2$

$H_o : p = 0.90$

(2) H_1 (สมมติฐานรอง, Alternative Hypothesis) หรือ

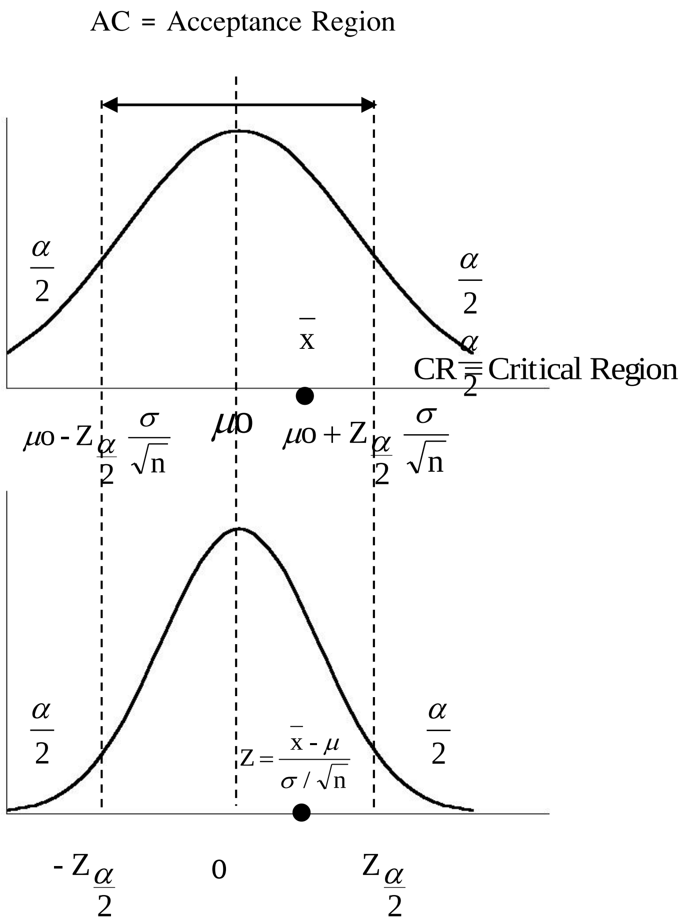
เช่น	$H_1 : \mu < 30{,}000$ กม	....	(One tail ด้านล่าง)
	$H_1 : \mu_1 \neq \mu_2$	....	(Two tails)
	$H_1 : p \neq 0.90$	....	(Two tails)

สมมติฐานรองคือเหตุการณ์ที่เป็น Complement ของสมมติฐานหลัก

การทดสอบแบบ Two Tails แสดงอยู่ในรูปที่ 5.1

$H_o : \mu = \mu_o$ (ค่าคงที่)

$H_1 : \mu \neq \mu_o$

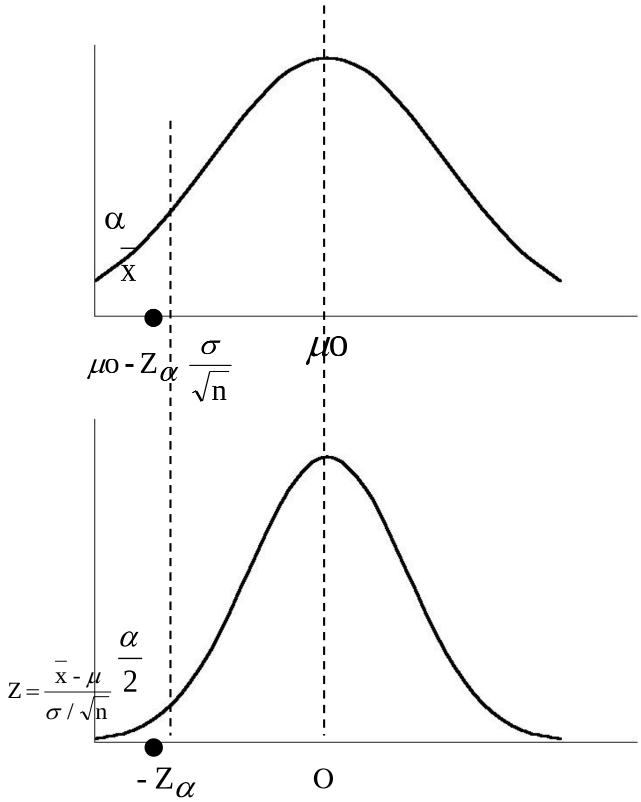


รูปที่ 5.1 การทดสอบสมมติฐานแบบ Two tails

การทดสอบแบบ One tail ด้านล่าง แสดงอยู่ในรูปที่ 5.2

$H_o : \mu \geq \mu_o$

$H_1 : \mu < \mu_o$

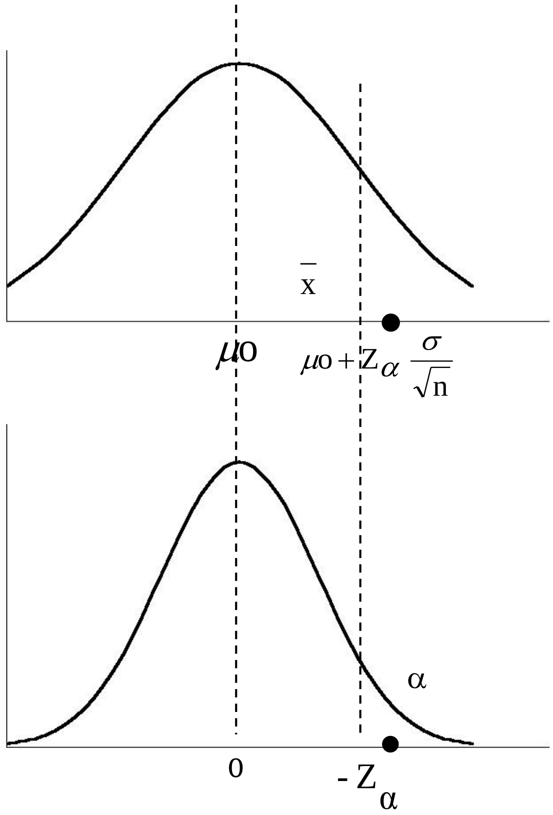


รูปที่ 5.2 การทดสอบสมมติฐานแบบ One tail ด้านล่าง

การทดสอบแบบ One tail ด้านบน แสดงอยู่ในรูปที่ 5.3

$H_o : \mu \leq \mu_o$

$H_1 : \mu > \mu_o$



รูปที่ 5.3 การทดสอบสมมติฐานแบบ One tail ด้านบน

5.4 ความผิดพลาดในการทดสอบ (Types of Error)

การทดสอบสมมติฐานอาจมีความผิดพลาดได้ 2 แบบคือ ความผิดพลาดประเภท I (Type I Error) และความผิดพลาดประเภท II (Type II Error)

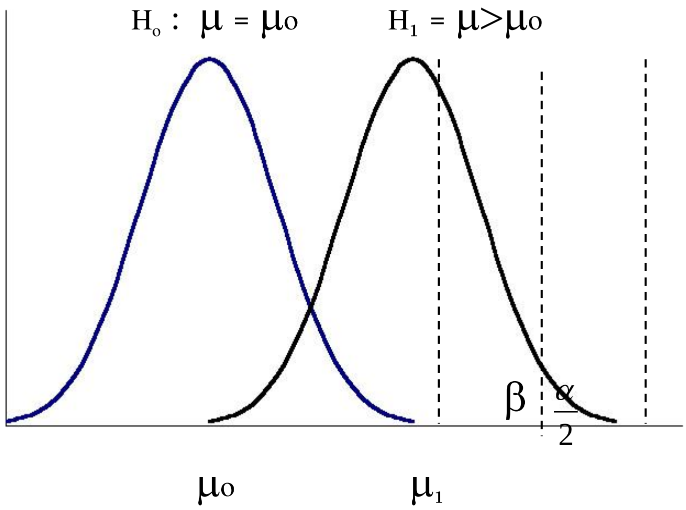
(1) ความผิดพลาดประเภท I หมายถึง Reject $H_o|H_o$ เป็นจริง

P(Type I Error) = α

(2) ความผิดพลาดประเภท II หมายถึง Accept $H_o|H_1$ เป็นจริง

P(Type II Error) = β

รูปที่ 5.4 แสดงการเกิดความผิดพลาดประเภท I และความผิดพลาดประเภท II ซึ่งสามารถสรุปความผิดพลาดในการทดสอบสมมติฐานได้ดังตารางที่ 5.1



α = Significant Level = P(Type I Error)

$(1-\alpha)$ = Acceptance Region ($-Z_{\frac{\alpha}{2}}$ to $+ Z_{\frac{\alpha}{2}}$ กรณีที่ตัวแปรสุ่มมีการแจกแจงแบบปกติ)

β = P(Type II Error)

$1-\beta$ = Power of The Test

รูปที่ 5.4 การเกิด Type I และ Type II Error

ตารางที่ 5.1 ข้อสรุปเกี่ยวกับการทดสอบสมมติฐาน

ข้อสรุปจากการทดสอบ	ข้อเท็จจริง	
	H_o จริง	H_1 จริง
ยอมรับ Ho	ตัดสินใจถูกต้อง $(1-\alpha)$	Type II Error β
ปฏิเสธ Ho	Type I Error α	ตัดสินใจถูกต้อง $(1-\beta)$

ตัวอย่างที่ 5.1 ในการทดสอบสมมติฐานว่านักเรียนทำข้อสอบปรนัยชนิดถูกผิดด้วยการเดาหรือไม่ ได้ตั้งข้อสมมติฐานไว้ดังนี้

H_0 : $p = 0.5$ (นักเรียนตอบคำถามโดยการเดา)

H_1 : $p \neq 0.5$ (นักเรียนตอบคำถามโดยไม่ได้เดา)

ทำการสุ่มตัวอย่าง คำตอบข้อสอบดังกล่าว 10 ข้อ และตั้งเกณฑ์ไว้ว่าถ้านักเรียนตอบถูกต้องตั้งแต่ 7 ข้อขึ้นไป แสดงว่านักเรียนไม่ได้เดา แต่ถ้าถูกน้อยกว่า 7 ข้อ แสดงว่านักเรียนเดา

จากเกณฑ์ดังกล่าว จงหาความผิดพลาดประเภท I และความผิดพลาดประเภท II ถ้าความจริงความน่าจะเป็นที่จะตอบคำถามแต่ละข้อได้ถูกต้องเป็น 0.7

ความผิดพลาดประเภท I $= \alpha$

$= P\,[\text{ปฏิเสธ } H_0 \mid H_0 \text{ เป็นจริง}]$

$= P\,[X \geq 7 \mid p = 0.5]$

$= 1-P\,[X < 7 \mid p = 0.5]$

$$= 1-\sum_{x=0}^{6} b\,(x\,;10, 0.5)$$

$= 1-0.8281$

$= 0.1719$

ความผิดพลาดประเภท II $= \beta$

$= P\,[\text{ยอมรับ } H_0 \mid H_1 \text{ เป็นจริง}]$

$= P\,[X < 7 \mid p = 0.7]$

$$= \sum_{x=0}^{6} b\,(x\,;10, 0.7)$$

$= 0.3504$

ตัวอย่างที่ 5.2 คาดว่าสัดส่วนของผู้จบการศึกษาคือ $p = 0.3$ และเพื่อทดสอบความจริงนี้ ได้มีการตั้งสมมติฐาน ดังนี้

H_0 : $p = 0.3$

H_1 : $p \neq 0.3$

ทำการสุ่มตัวอย่างนิสิต 15 คน และตั้งเกณฑ์ไว้ว่าถ้านิสิตเหล่านี้จบการศึกษา 2 ถึง 7 คน จะยอมรับสมมติฐาน H_0 แต่ถ้าได้ผลอย่างอื่นจะสรุปว่า H_1 เป็นจริง จงหาความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ I และ II เมื่อ p = 0.2

ความผิดพลาดประเภท I $= \alpha$

$= P$ [ปฏิเสธ H_0 | H_0 เป็นจริง]

$= P[X < 2, X > 7 \mid p = 0.3]$

$= 1 - \sum_{x=2}^{7} b(x;15,0.3)$

$= 1 - [\sum_{x=0}^{7} b(x;15,0.3) - \sum_{x=0}^{1} b(x;15,0.3)]$

$= 1 - [0.95 - 0.0353]$

$= 0.0853$

ความผิดพลาดประเภท II $= \beta$

$= P$ [ยอมรับ H_0 | H_1 เป็นจริง]

$= P[2 \leq X \leq 7 \mid p = 0.2]$

$= \sum_{x=2}^{7} b(x;15,0.2) - \sum_{x=0}^{1} b(x;15,0.2)$

$= 0.9958 - 0.1671$

$= 0.8287$

คุณสมบัติเกี่ยวกับค่าความผิดพลาด (α,β) พอสรุปได้ดังนี้

(1) เพิ่ม Acceptance Region ของ H_0 หรือเพิ่มค่า (1-α)

จะลด α = P (Type I Error)

แต่จะเพิ่ม β = P (Type II Error)

ยิ่ง H_0 ใกล้ H_1 จะยิ่งเพิ่ม β มากขึ้น

(2) เพิ่ม n จะทำให้ α และ β ลดลง

(3) ถ้า H_1 เป็นจริง ค่า β จะลดลงเมื่อค่าจริงต่างจากค่าตาม Ho มาก

ดังนั้น ถ้าต้องการลดค่าความผิดพลาดจะทำได้โดยการเพิ่มค่า n

ตัวอย่างที่ 5.3 เหมืองแร่แห่งหนึ่งอ้างว่าแร่เหล็กจากเหมืองแห่งนี้มีเปอร์เซ็นต์เหล็กในเนื้อแร่เฉลี่ยร้อยละ 60 และมีส่วนเบี่ยงเบนมาตรฐาน 10% เพื่อทดสอบข้ออ้างนี้ จึงได้ตั้งสมมติฐานว่า

$H_0 \quad : \quad \mu \leq 60\%$

แย้งกับ $H_1 \quad : \quad \mu > 60\%$

และได้มีการสุ่มตัวอย่างแร่เหล็กเพื่อทดสอบสมมติฐานดังกล่าวโดยตั้งเกณฑ์ว่าจะยอมรับ H_0 เมื่อค่าเฉลี่ยของเปอร์เซ็นต์เหล็กในเนื้อแร่ ไม่เกินร้อยละ 62 มิฉะนั้น จะยอมรับ H_1

จงหาความผิดพลาดประเภท I (α) และประเภท II (β) เมื่อแท้จริงแล้วค่าเฉลี่ยของเปอร์เซ็นต์เหล็กในเนื้อแร่เท่ากับ 65%

ก. เมื่อทำการสุ่มตัวอย่าง 25 ตัวอย่าง

ข. เมื่อทำการสุ่มตัวอย่าง 100 ตัวอย่าง

ก. α = P [ปฏิเสธ $H_0 \mid H_0$ เป็นจริง]

$= P(\overline{X} > 62 \mid \mu = 60\%)$

$= P(Z > \frac{62 - 60}{10/\sqrt{25}})$

$= 0.1587$

β = P [ยอมรับ $H_0 \mid H_1$ เป็นจริง]

$= P(\overline{X} < 62 \mid \mu = 65\%)$

$= P(Z < \frac{62 - 60}{10/\sqrt{25}})$

$= 0.0668$

ข. $\alpha = P(Z > \frac{62 - 60}{10/\sqrt{100}})$

$= 0.0228$

$\beta = P(Z < \frac{62 - 65}{10/\sqrt{100}})$

$= 0.0013$

ดังนั้นจะเห็นได้ว่าเมื่อเพิ่มขนาดของตัวอย่างทำให้ค่าความผิดพลาด α และ β ลดลงทั้งคู่

5.5 การทดสอบสมมติฐานของ μ

5.5.1 กรณีรู้ σ^2 ใช้ $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

H_o	H_1	Critical Region
$\mu = \mu_o$	$\mu \neq \mu_o$	$\lvert z\rvert = \left\lvert\dfrac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}\right\rvert > z_{\frac{\alpha}{2}}$
$\mu \geq \mu_o$	$\mu < \mu_o$	$z = \dfrac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} < -z_\alpha$
$\mu \leq \mu_o$	$\mu > \mu_o$	$z = \dfrac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} > z_\alpha$

ตัวอย่างที่ 5.4 ขนมที่ผลิตจากโรงงานแห่งหนึ่งมีน้ำหนักเฉลี่ยกล่องละ 8 ออนซ์ และมีค่าส่วนเบี่ยงเบนมาตรฐาน 0.16 ออนซ์ ในการทดสอบว่าในวันหนึ่ง ๆ การผลิตเป็นไปตามทีกำหนดไว้หรือไม่ ได้ทำการสุ่มตัวอย่างขนมที่ผลิตในแต่ละวันจำนวน 25 ชิ้น พบว่ามีน้ำหนักเฉลี่ย 8.112 ออนซ์ เนื่องจากโรงงานต้องเสียค่าใช้จ่ายเพิ่มขึ้นถ้าน้ำหนักเกิน 8 ออนซ์ และลูกค้าต้องเสียเงินซื้อขนมแพงขึ้น ถ้าน้ำหนักไม่ถึง 8 ออนซ์ ดังนั้นจึงกำหนดสมมติฐานรองคือ $\mu \neq 8$ ที่ระดับนัย $\alpha = 0.01$ จงทดสอบว่าการผลิตเป็นไปตามที่กำหนดหรือไม่

H_o : $\mu = 8$

แย้งกับ H_1 : $\mu \neq 8$

ระดับนัยสำคัญ $\alpha = 0.01$

ช่วงวิกฤต $z > z_{0.005}$ หรือ $z > 2.58$

$z < -z_{0.005}$ หรือ $z < -2.58$

ค่าสถิติจากการสุ่มตัวอย่างขนาด 25 ตัวอย่าง และ $\bar{x} = 8.112$ ออนซ์

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{8.112 - 8.0}{0.16/\sqrt{25}}$$

$$z = 3.5$$

สรุปผล : เนื่องจาก z อยู่ในบริเวณวิกฤต จึงปฏิเสธ H_o ซึ่งหมายความว่าการผลิตไม่เป็นไปตามที่กำหนดไว้ และต้องมีการปรับปรุงการผลิตใหม่

5.5.2 กรณีไม่รู้ σ^2 แต่ n > 30

เหมือนกรณี 8.5.1 แต่ $z = \dfrac{\bar{x} - \mu_o}{S/\sqrt{n}}$

5.5.3 กรณีไม่รู้ σ^2 แต่ n < 30

ใช้ $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$; $\upsilon = n - 1$

H_o	H_1	Critical Region
$\mu = \mu_o$	$\mu \neq \mu_o$	$\lvert t\rvert = \left\lvert\dfrac{\bar{x} - \mu_o}{S/\sqrt{n}}\right\rvert > t_{\frac{\alpha}{2},\upsilon}$
$\mu \geq \mu_o$	$\mu < \mu_o$	$t = \dfrac{\bar{x} - \mu_o}{S/\sqrt{n}} < -t_{\alpha,\upsilon}$
$\mu \leq \mu_o$	$\mu > \mu_o$	$t = \dfrac{\bar{x} - \mu_o}{S/\sqrt{n}} > t_{\alpha,\upsilon}$

ตัวอย่างที่ 5.5 สมมติว่าผ้าที่ผลิตจากโรงงานแห่งหนึ่ง สามารถทนต่อแรงดึงได้ 185 ปอนด์ ในการทดสอบผ้าที่ผลิตได้ว่ามีคุณสมบัติตามที่กำหนดหรือไม่ ได้สุ่มตัวอย่างผ้าจำนวน 5 ชิ้น เพื่อทดสอบแรงดึง พบว่าผ้าสามารถทนแรงดึงเฉลี่ย 183.1 ปอนด์ และมีค่าส่วนเบี่ยงเบนมาตรฐาน 8.2 ปอนด์ จงทดสอบสมมติฐานดังกล่าวด้วยระดับนัยสำคัญ 0.05

H_o : $\mu = 185$

แย้งกับ H_1 : $\mu < 185$

ระดับนัยสำคัญ $\alpha = 0.05$

ช่วงวิกฤต $t < -t_{0.05,4}$

$t < -2.132$

ค่าสถิติที่ได้จากการสุ่มตัวอย่าง 5 ตัวอย่าง และ $\bar{x} = 183.1$, $s = 8.2$

$$T = \frac{183.1 - 185}{8.2/\sqrt{5}}$$

$$T = -0.49$$

สรุปผล : เนื่องจาก t ไม่อยู่ในช่วงวิกฤต ดังนั้นจึงไม่สามารถปฏิเสธ H_o ได้หรือจำเป็นต้องยอมรับว่าผ้าที่ผลิตจากโรงงานนี้สามารถทนต่อแรงดึงได้ 185 ปอนด์ จริงที่ระดับนัยสำคัญ 0.05

5.6 การทดสอบสมมติฐานของ $\mu_1-\mu_2$

5.6.1 กรณีทราบ σ_1^2, σ_2^2

ใช้ $$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

H_o	H_1	Critical Region
$\mu_1 - \mu_2 = 0$	$\mu_1 \neq \mu_2$	$\lvert z\rvert > Z_{\frac{\alpha}{2}}$
	$\mu_1 < \mu_2$	$z < -Z_\alpha$
	$\mu_1 > \mu_2$	$z > Z_\alpha$
$\mu_1 - \mu_2 = d$	$\mu_1 - \mu_2 \neq d$	$\lvert z\rvert > Z_{\frac{\alpha}{2}}$
	$\mu_1 - \mu_2 < d$	$z < -Z_\alpha$
	$\mu_1 - \mu_2 > d$	$z > Z_\alpha$

ตัวอย่างที่ 5.6 สารนิโคตินในบุหรี่ 2 ชนิด ซึ่งมีค่าส่วนเบี่ยงเบนมาตรฐาน $\sigma_1 = 1.2$ มิลลิกรัม และ $\sigma_2 = 1.4$ มิลลิกรัม ผู้ผลิตอ้างว่าสารนิโคตินในบุหรี่ชนิดแรกมากกว่าชนิดที่สองอยู่ 2 มิลลิกรัม ในการทดสอบข้อกล่าวอ้างดังกล่าว ได้ทำการสุ่มตัวอย่างบุหรี่ชนิดแรก 50 มวน พบว่ามีค่าเฉลี่ยของปริมาณสารนิโคติน 26.1 มิลลิกรัม และสุ่มตัวอย่างบุหรี่ชนิดที่สอง 40 มวน พบว่ามีค่าเฉลี่ยของปริมาณสารนิโคติน 23.8 มิลลิกรัม จงทดสอบสมมติฐานว่าสารนิโคตินในบุหรี่ชนิดแรงมากกว่าชนิดที่สองอยู่ 2 มิลลิกรัมตามข้อกล่าวอ้างที่ระดับนัยสำคัญ 0.05

$$H_o : \mu_1-\mu_2 = 2.0$$

แย้งกับ H_1 : $\mu_1-\mu_2 \neq 2.0$

ระดับนัยสำคัญ α = 0.05

ช่วงวิกฤต $z > z_{0.025}$ หรือ $z > 1.96$

$z < -z_{0.025}$ หรือ $z < -1.96$

ค่าสถิติที่ได้จากการสุ่มตัวอย่าง

$n_1 = 50$, $\bar{x}_1 = 26.1$ มิลลิกรัม

$n_2 = 40$, $\bar{x}_2 = 23.8$ มิลลิกรัม

$$z = \frac{(26.1 - 23.8) - 2.0}{\sqrt{1.2^2/50 + 1.4^2/40}}$$

$$= 1.08$$

สรุปผล : เนื่องจาก z ไม่อยู่ในวิกฤต จึงไม่สามารถปฏิเสธ H_o ได้ที่ระดับนัยสำคัญ 0.05 ดังนั้นจึงสรุปว่า $\mu_1-\mu_2 = 2.0$ มิลลิกรัม

5.6.2 กรณีไม่ทราบ σ_1^2 , σ_2^2 แต่ $n_1, n_2 > 30$

เหมือนกรณี 5.6.1 แต่ $$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

5.6.3 กรณีไม่ทราบ σ_1^2 , σ_2^2 และ $n_1, n_2 < 30$

(1) กรณี $\sigma_1^2 = \sigma_2^2$

ใช้ $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$Sp = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 - n_2 - 2}} \; ; \; \upsilon = n_1 - n_2 - 2$$

Ho	H_1	Critical Region
$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$\lvert t \rvert > t_{\frac{\alpha}{2},\upsilon}$
	$\mu_1 - \mu_2 < 0$	$t < -t_{\alpha,\upsilon}$
	$\mu_1 - \mu_2 > 0$	$t > t_{\alpha,\upsilon}$
$\mu_1 - \mu_2 = d$	$\mu_1 - \mu_2 \neq d$	$\lvert t \rvert > t_{\frac{\alpha}{2},\upsilon}$
	$\mu_1 - \mu_2 < d$	$t < -t_{\alpha,\upsilon}$
	$\mu_1 - \mu_2 > d$	$t > t_{\alpha,\upsilon}$

ตัวอย่างที่ 5.7 ในการเปรียบเทียบสีทาบ้าน 2 ชนิด ได้ทำการสุ่มตัวอย่างสีชนิดถังละ 1 แกลลอน จากสีทั้งสองชนิด ชนิดละ 4 ถัง เพื่อทดสอบ ปรากฏว่า สีชนิดแรกทาได้เฉลี่ยถังละ 512 ตารางฟุต โดยมีส่วนเบี่ยงเบนมาตรฐาน 31 ตารางฟุต และสีชนิดที่สองทาได้เฉลี่ยถังละ 492 ตารางฟุต โดยมีส่วนเบี่ยงเบนมาตรฐาน 26 ตารางฟุต จงทดสอบสมมติฐานว่าสีทั้งสองชนิดสามารถทาได้พื้นที่เฉลี่ยเท่ากันที่ระดับนัยสำคัญ 0.05 ถ้าทราบว่าค่าความแปรปรวนของพื้นที่ของสีทั้งสองเท่ากัน

H_o : $\mu_1 - \mu_2$

แย้งกับ H_1 : $\mu_1 \neq \mu_2$

ระดับนัยสำคัญ α = 0.05

υ = 4+4-2 = 6

ช่วงวิกฤต $t > t_{\alpha/2}\upsilon$ หรือ $t > 2.447$

$t < -t_{\alpha/2}\upsilon$ หรือ $t < -2.447$

ค่าสถิติที่ได้จากการสุ่มตัวอย่าง

$n_1 = 4$, $\bar{x}_1 = 512$, $S_1 = 31$

$n_2 = 4$, $\bar{x}_2 = 492$, $S_2 = 26$

$$t = \frac{(512 - 492) - 0}{\sqrt{\frac{3(31)^2 + 3(26)^2}{6}}\sqrt{\frac{1}{4} + \frac{1}{4}}} = 0.9886$$

สรุปผล : เนื่องจาก t ไม่อยู่ในช่วงวิกฤต ดังนั้นจึงไม่สามารถปฏิเสธ H_o ได้หรือจำเป็นต้องยอมรับว่าสีทั้งสองสามารถทาได้พื้นที่เฉลี่ยเท่ากันที่ระดับนัยสำคัญ 0.05

(2) กรณี $\sigma_1^2 \neq \sigma_2^2$

เหมือนกรณี 5.6.3 (1) แต่

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$\upsilon = \frac{[S_1^2/n_1 + S_2^2/n_2]^2}{(\frac{S_1^2}{n_1})^2(\frac{1}{n_1 - 1}) + (\frac{S_2^2}{n_2})^2(\frac{1}{n_{2-1}})}$$

ตัวอย่างที่ 5.8 ในการทดสอบว่าวัสดุชนิดที่ 1 สามารถรับน้ำหนักได้มากกว่าวัสดุชนิดที่สองหรือไม่ ได้มีการสุ่มตัวอย่างวัสดุชนิดที่ 1 จำนวน 25 ชิ้น และวัสดุชนิดที่ 2 จำนวน 16 ชิ้น เพื่อทดสอบการรับน้ำหนักปรากฏผลดังนี้

วัสดุชนิดที่ 1				วัสดุชนิดที่ 2			
n_1	=	25	ชิ้น	n_2	=	16	ชิ้น
$\bar{x}_1$	=	380	ปอนด์	$\bar{x}_2$	=	370	ปอนด์
S_1^2	=	100	ปอนด์	S_2^2	=	400	ปอนด์

จงทดสอบสมมติฐานที่ว่าวัสดุชนิดที่ 1 สามารถรับน้ำหนักเฉลี่ยได้มากกว่าวัสดุชนิดที่ 2 ที่ระดับนัยสำคัญ 0.05 สมมติว่าค่าความแปรปรวนของการรับน้ำหนักของวัสดุทั้งสองไม่เท่ากัน

H_o : $\mu_1 = \mu_2$

แย้งกับ H_1 : $\mu_1 > \mu_2$

ระดับนัยสำคัญ α = 0.05

ช่วงวิกฤต $t > t_{\alpha,\upsilon}$

$$\upsilon = \frac{[100/25 + 400/16]^2}{\frac{(100/25)^2}{24} + \frac{(400/16)^2}{15}}$$

$= \quad 19.86$

หรือองศาเสรี $= \quad 20$

ช่วงวิกฤต $t > t_{0.05,\,20}$ หรือ $t > 1.725$

ค่าสถิติที่ได้จากการสุ่มตัวอย่าง

$$t = \frac{(380 - 370) - 0}{\sqrt{100/25 - 400/16}}$$

$$= 1.857$$

สรุปผล : เนื่องจาก t อยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.05 หรือสรุปว่าวัสดุชนิดที่ 1 สามารถทนน้ำหนักได้มากกว่าวัสดุชนิดที่ 2

5.7 การทดสอบสมมติฐานของ μ_d

กำหนดให้ x_I และ y_I เป็นค่าที่สุ่มเป็นคู่

$$d_i = x_i - y_i$$

ใช้

$$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}}, \upsilon = n - 1$$

H_o	H_1	Critical Region
$\mu_d = d_o$	$\mu_d \neq d_0$	$\lvert t\rvert > t_{\frac{\alpha}{2},\upsilon}$
	$\mu_d < d_0$	$t < -t_{\alpha,\upsilon}$
	$\mu_d > d_0$	$t > t_{\alpha,\upsilon}$

$$t = \frac{\bar{d} - d_o}{S_d / \sqrt{n}}$$

ตัวอย่างที่ 5.9 ผู้คิดโปรแกรมคอมพิวเตอร์ในการคำนวณโจทย์คณิตศาสตร์ 2 โปรแกรม ต้องการทราบว่าโปรแกรมทั้งสองใช้เวลาในการคำนวณแตกต่างกันหรือไม่ จึงได้มีการสุ่มตัวอย่างโจทย์จำนวน 10 ข้อ และใช้โปรแกรมทั้งสองคำนวณได้ผลดังนี้

โจทย์ข้อที่	1	2	3	4	5	6	7	8	9	10
โปรแกรมเก่า (วินาที)	8.05	24.74	28.33	8.45	9.19	25.10	14.05	20.33	4.82	8.54
โปรแกรมใหม่ (วินาที)	0.71	0.74	0.74	0.77	0.80	0.83	0.82	0.77	0.71	0.72

จงทดสอบสมมติฐานว่าโปรแกรมใหม่สามารถคำนวณได้รวดเร็วกว่าโปรแกรมเก่าที่ระดับนัยสำคัญ 0.0005

$H_o \quad : \quad \mu_d \quad = \quad 0$

$H_1 \quad : \quad \mu_d \quad > \quad 0$

ระดับนัยสำคัญ $\alpha = 0.0005$

ช่วงวิกฤต $t > t_{0.0005,9}$ หรือ $t > 4.781$

ค่าสถิติจากการสุ่มตัวอย่างคู่จำนวน 10 คู่ คำนวณหาค่า $\bar{d}$, S_d

โจทย์ข้อที่	1	2	3	4	5	6	7	8	9	10
d (วินาที)	7.34	24.00	27.59	7.68	8.39	24.37	13.23	19.56	4.11	7.82

$$\bar{d} = \frac{\sum_{i=1}^{10} d_i}{10} = 14.41$$

$$S_d = 8.65$$

$$t = \frac{14.41 - 0}{8.65/\sqrt{10}}$$

$$t = 5.268$$

สรุปผล : เนื่องจากค่า t อยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.0005 และสรุปว่าการคำนวณโดยใช้โปรแกรมใหม่จะใช้เวลาน้อยกว่าโปรแกรมเก่า

5.8 การทดสอบสมมติฐานของ p

ใช้ $z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}}}$

H_o	H_1	Critical Region
$p = p_o$	$p \neq p_o$	$\lvert z\rvert > Z_{\frac{\alpha}{2}}$
	$p < p_o$	$z < -Z_{\alpha}$
	$p > p_o$	$z > Z_{\alpha}$

$$z = \frac{\hat{p}-p_o}{\sqrt{\dfrac{p_o q_o}{n}}}$$

ตัวอย่างที่ 5.10 ได้มีการกล่าวอ้างว่าประชาชนในเมืองเมืองหนึ่งมีความนิยมในตัวนักการเมืองผู้หนึ่ง จำนวนร้อยละ 25 ในการทดสอบข้อกล่าวอ้างนี้ได้ทำการสุ่มตัวอย่างประชากรในเมืองนั้น ๆ จำนวน 100 คน เพื่อสัมภาษณ์และพบว่าชอบนักการเมืองผู้นั้นจำนวน 20 คน จงสรุปผลที่ระดับนัยสำคัญ 0.05 ว่าคำกล่าวอ้างนั้นถูกต้องหรือไม่

H_o : $p = 0.25$

แย้งกับ H_1 : $p \neq 0.25$

ระดับนัยสำคัญ α = 0.05

ช่วงวิกฤต $z > z_{\alpha/2}$ หรือ $z > 1.96$

$z < -z_{\alpha/2}$ หรือ $z < -1.96$

ค่าสถิติ

$$z = \frac{\hat{p}-p_o}{\sqrt{p_o q_o / n}} = \frac{20/100 - 0.25}{\sqrt{(0.25x0.75)/100}} = -1.15$$

6. สรุปผล : เนื่องจาก z อยู่นอกบริเวณวิกฤต ดังนั้นจึงไม่สามารถปฏิเสธ H_o ได้ที่ระดับนัยสำคัญ 0.05 หรือสรุปได้ว่าประชาชนในเมืองนี้มีความนิยมนักการเมืองผู้นี้จำนวนร้อยละ 25 จริง

5.9 การทดสอบสมมติฐานของ p_1-p_2 กรณีทราบ p_1 และ p_2

ใช้ $$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

H_o	H_1	Critical Region
$p_1-p_2 = 0$	$p_1-p_2 \neq 0$	$\lvert z\rvert > z_{\frac{\alpha}{2}}$
	$p_1-p_2 < 0$	$z < -z_{\alpha}$
	$p_1-p_2 > 0$	$z > z_{\alpha}$

กรณีไม่ทราบ p_1, p_2 แต่ $p_1 = p_2$ และ $n_1, n_2 > 30$

$$\hat{p}_1 = \frac{x_1}{n_1}$$

$$\hat{p}_2 = \frac{x_2}{n_2}$$

ให้ $\hat{p} \sim p_1 = p_2$ และ $\hat{q} \sim q_1 = q_2$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \; ; \; \hat{q} = 1 - \hat{p}$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

ตัวอย่างที่ 5.11 ในการทดสอบน้ำยาซักผ้าสองชนิด พบว่าน้ำยาซักผ้าชนิดแรกสามารถขจัดความสกปรกออกได้ 63 ครั้ง ในการทดลอง 91 ครั้ง และชนิดที่สองสามารถขจัดความสกปรกออกได้ 42 ครั้ง ในการทดลอง 79 ครั้ง จากผลการทดสอบดังกล่าวจะสรุปได้หรือไม่ว่าน้ำยาซักผ้าทั้งสองมีประสิทธิภาพในการขจัดความสกปรกได้เท่ากัน ที่ระดับนัยสำคัญ 0.10

$$H_o \quad : \quad p_1 = p_2$$

แย้งกับ H_1 : $p_1 \neq p_2$

ระดับนัยสำคัญ α = 0.10

ช่วงวิกฤต $z > 1.645$

$z < -1.645$

ค่าสถิติจากการสุ่มตัวอย่าง $n_1 = 91$ $\hat{p}_1 = \frac{63}{91} = 0.692$

$n_2 = 79$ $\hat{p}_2 = \frac{42}{79} = 0.532$

$$\hat{p} = \frac{63+42}{91+79} = 0.618$$

$$\hat{q} = 1-0.618 = 0.382$$

$$z = \frac{0.692 - 0.532}{\sqrt{0.618 \times 0.382\,(1/91 + 1/79)}}$$

$$= 2.166$$

สรุปผล : เนื่องจากค่า z อยู่ในช่วงวิกฤต ดังนั้นจึงปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.10 และสรุปได้ว่าน้ำยาซักผ้าทั้งสองชนิดมีประสิทธิภาพไม่เท่ากัน

5.10 การทดสอบสมมติฐานของ σ^2

ใช้ $\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}$, $\upsilon = n-1$

H_o	H_1	Critical Region
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\frac{\alpha}{2},\upsilon}$ และ $\chi^2 > \chi^2_{\frac{\alpha}{2},\upsilon}$
	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha,\upsilon}$
	$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha,\upsilon}$

ตัวอย่างที่ 5.12 ศาสตราจารย์ด้านจิตวิทยาผู้หนึ่งกล่าวว่าไอคิวของนักศึกษามีค่าความแปรปรวนเท่ากับ 100 คะแนน ในการทดสอบข้อกล่าวอ้างดังกล่าวได้ทำการสุ่มตัวอย่างนักศึกษาจำนวน 23 คน เพื่อทดสอบไอคิวและหาค่าความแปรปรวน พบว่าค่าความแปรปรวนของนักศึกษากลุ่มนี้เท่ากับ 147.82 คะแนน จงทดสอบสมมติฐานที่ว่าความแปรปรวนของไอคิวนักศึกษายังคงเท่ากับ 100 คะแนนหรือไม่ ที่ระดับนัยสำคัญ 0.05

$H_o \quad : \quad \sigma^2 = 100$

แย้งกับ $H_1 \quad : \quad \sigma^2 \neq 100$

ระดับนัยสำคัญ $\alpha = 0.05$

ช่วงวิกฤต $\chi^2 > \chi^2_{0.025,\,22}$

$\chi^2 > 36.78$

และ $\chi^2 < \chi^2_{0.975,\,22}$

$\chi^2 < 10.98$

ค่าสถิติจากการสุ่มตัวอย่าง

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$\chi^2 = \frac{22 \text{ x } 147.82}{100}$$

$$\chi^2 = 32.52$$

สรุปผล : เนื่องจากค่าสถิติที่คำนวณจากตัวอย่าง อยู่นอกช่วงวิกฤต ดังนั้นจึงไม่สามารถปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.05 นั่นคือไอคิวของนักศึกษา ยังคงมีความแปรปรวนเท่ากับ 100 คะแนน

5.11 การทดสอบสมมติฐานของ σ_1^2, σ_2^2

ใช้ $f = \dfrac{S_1^2}{S_2^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2}$ $(\upsilon_1 = n_1 - 1 \,;\, \upsilon_2 = n_2 - 1)$

H_o	H_1	Critical Region
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$f < f_{1-\frac{\alpha}{2}, \upsilon_1, \upsilon_2}$ และ $f > f_{\frac{\alpha}{2}, \upsilon_1, \upsilon_2}$
	$\sigma_1^2 < \sigma_2^2$	$f < f_{1-\alpha, \upsilon_1, \upsilon_2}$
	$\sigma_1^2 > \sigma_2^2$	$f > f_{\alpha, \upsilon_1, \upsilon_2}$

ตัวอย่างที่ 5.13 ในการทดสอบเพื่อเปรียบเทียบค่าความแปรปรวนของ Tensile Strength ของเหล็กโครงสร้าง 2 ชนิด ได้ทำการสุ่มตัวอย่างเหล็กโครงสร้างชนิดแรก 13 ตัวอย่าง หาค่าความแปรปรวนได้ 19.2 x 10^3 ปอนด์ต่อตารางนิ้ว และสุ่มตัวอย่างเหล็กโครงสร้างชนิดที่สอง 16 ตัวอย่าง หาค่าความแปรปรวนได้ 3.5 x 10^3 ปอนด์ต่อตารางนิ้ว จงทดสอบสมมติฐานว่าเหล็กโครงสร้างทั้งสองชนิดมีความแปรปรวนของ Tensile Strength เท่ากันหรือไม่ ที่ระดับนัยสำคัญ 0.02

Ho : $\sigma_1^2 = \sigma_2^2$

แย้งกับ H_1 : $\sigma_1^2 \neq \sigma_2^2$

ระดับนัยสำคัญ α = 0.02

ช่วงวิกฤต $f > f_{0.01(12, 15)}$

$f > 3.67$

และ $f < f_{0.99\,(12,15)}$

$$f < \frac{1}{f_{0.01(15,12)}}$$

$$f < \frac{1}{4.01}$$

$f < 0.25$

ค่าสถิติจากการสุ่มตัวอย่าง

$f = S_1^2 / S_2^2$

$$f = \frac{19.2}{3.5}$$

$f = 5.49$

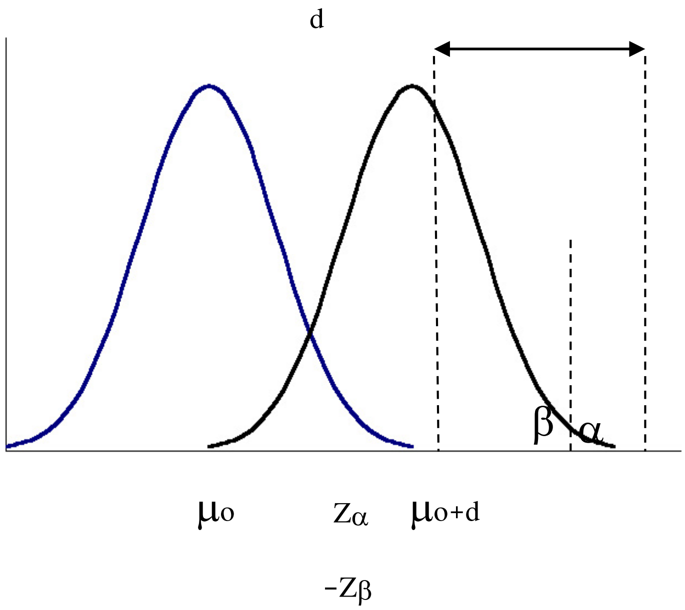
สรุปผล : เนื่องจากค่าสถิติ f อยู่ในช่วงวิกฤตดังนั้นจึงปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.02 นั่นคือค่าความแปรปรวนของ Tensile Strength ของเหล็กโครงสร้างทั้งสองชนิดไม่เท่ากัน

5.12 การหาขนาดตัวอย่าง n ในการทดสอบ μ เมื่อทราบ σ^2

จากรูปที่ 5.5 α = P (Type I Error)

และ β = P (Type II Error)

ถ้า n เพิ่ม จะทำให้ α, β ลดลง



รูปที่ 5.5 ความผิดพลาดประเภท I และ II ในการทดสอบ μ

ถ้ากำหนด α และ β จะหา n ได้ดังนี้

(1) กรณี One Tail ด้านบน

H_o : $\mu \le \mu_o$

H_1 : $\mu > \mu_o$ หรือ $\mu_1 = \mu_{o+d}$

$$\beta = P[\text{Accept } H_o \mid H_1 \text{ เป็นจริง}]$$

$$= P(Z < z_\alpha \mid \mu = \mu_{o+d})$$

$$= P\left[\frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} < z_\alpha \mid \mu = \mu_{o+d}\right]$$

$$= P\left[\frac{\bar{x} - (\mu_o + d)}{\sigma/\sqrt{n}} < z_\alpha - \frac{d}{\sigma/\sqrt{n}}\right]$$

$$P(Z < -z_\beta) = P\left[\frac{\bar{x} - (\mu_o + d)}{\sigma/\sqrt{n}} < z_\alpha - \frac{d}{\sigma/\sqrt{n}}\right]$$

$$-z_\beta = z_\alpha - \frac{d}{\sigma/\sqrt{n}}$$

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{d^2}$$

(2) กรณี One Tail ด้านล่าง

$H_o \quad : \quad \mu = \mu o$

$H_1 \quad : \quad \mu < \mu o$ หรือ $\mu_1 = \mu o\text{-}d$

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{d^2}$$

(3) กรณี Two Tails

$H_o \quad : \quad \mu = \mu o$

$H_1 \quad : \quad \mu \neq \mu o$

$$n = \frac{(z_{\frac{\alpha}{2}} + z_{\beta})^2 \sigma^2}{d^2}$$

ตัวอย่างที่ 5.14 ในการทดสอบสมมติฐานว่าค่าเฉลี่ยของประชากรมีค่าเกินกว่า 28 หรือไม่ ที่ระดับนัยสำคัญ 0.05 ถ้ากำหนดอำนาจการทดสอบเป็น 0.99 และทราบว่าค่าความแปรปรวนของประชากรเท่ากับ 12 และค่าเฉลี่ยที่แท้จริงของประชากรเท่ากับ 32 จงหาขนาดตัวอย่างที่เหมาะสม

$H_o \quad : \quad \mu = 28$

แย้งกับ $H_1 \quad : \quad \mu > 28$

$\mu_1 = 28+4,\ d = 4$

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{d^2}$$

$$n = \frac{(1.645 + 2.325)^2 \text{ x } 12^2}{4 \text{ x } 4}$$

$n = 141.85$

ดังนั้นจะต้องใช้ตัวอย่างจำนวน 142 ตัวอย่างในการที่จะปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.05 และยอมรับ H_1 ด้วยอำนาจการทดสอบ 0.99 โดยที่ค่าเฉลี่ยที่แท้จริงคือ 32

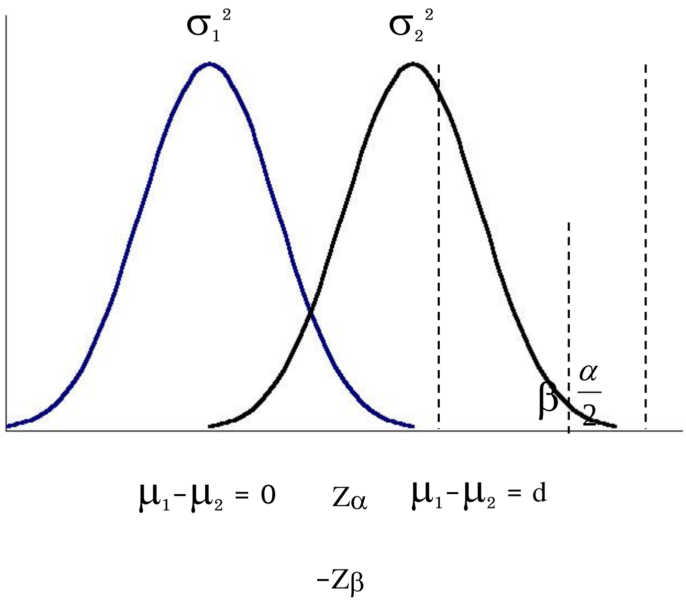
5.13 การหาขนาดตัวอย่าง n ในการทดสอบ μ_1, μ_2 กรณีทราบ σ_1^2, σ_2^2

ให้ $n_1 = n_2 = n$

(1) กรณี Two Tails (ดูรูปที่ 5.6)

H_o : $\mu_1 - \mu_2 = 0$

H_1 : $\mu_1 - \mu_2 \neq 0$ หรือ $\mu_1 - \mu_2 = d$



รูปที่ 5.6 ความผิดพลาดประเภท I และ II ในการทดสอบ μ_1-μ_2

$$\beta = P[\text{ Type II Error}]$$

$$= P[\text{Accept Ho} \mid H_1 \text{ เป็นจริง}]$$

$$= P[-z_{\frac{\alpha}{2}} \le Z < z_{\frac{\alpha}{2}} \mid \mu_1 - \mu_2 = d]$$

$$= P\left[-z_{\frac{\alpha}{2}} \le \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\frac{\alpha}{2}} \mid \mu_1 - \mu_2 = d\right]$$

$$\beta = P\left[\frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\frac{\alpha}{2}} \mid \mu_1 - \mu_2 = d\right]$$

$$P(Z < -z_\beta) = P\left[\frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\frac{\alpha}{2}} - \frac{d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right]$$

$$-Z_\beta = z_{\frac{\alpha}{2}} - \frac{d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$n = \frac{(z_{\frac{\alpha}{2}} + z_\beta)^2 (\sigma_1^2 + \sigma_2^2)}{d^2}$$

(2) กรณี One Tail

$$n = \frac{(z_\alpha + z_\beta)^2 (\sigma_1^2 + \sigma_2^2)}{d^2}$$

ตัวอย่างที่ 5.15 สมมติว่าขนาดของตัวอย่าง n สุ่มจากประชากรปกติ 2 ชุด ซึ่งทราบค่าความแปรปรวน $\sigma_1^2 = 12$ และ $\sigma_2^2 = 15$ เพื่อทดสอบสมมติฐานหลัก

H_o : $\mu_1 - \mu_2 = 110$ แย้งกับข้อสมมติฐานรอง

H_1 : $\mu_1 - \mu_2 \neq 110$ และถ้า $\mu_1 - \mu_2 = 113$

ความน่าจะเป็นของการเกิดความผิดพลาดประเภท I และประเภท II ที่กำหนดให้เท่ากับ 0.01 จงหาขนาดของตัวอย่างที่เหมาะสม

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{d^2}$$

$$n = \frac{(2.575 + 2.325)^2 (12^2 + 15^2)}{(110 - 113)^2}$$

$$n = 894.17$$

ดังนั้นจะต้องใช้ตัวอย่างจำนวน 895 ตัวอย่างในการปฏิเสธ H_o ที่ระดับนัยสำคัญ 0.01 และยอมรับ H_1 ด้วยอำนาจการทดสอบ 0.99 และผลต่างของค่าเฉลี่ยที่แท้จริงคือ 113

5.14 การหาขนาดตัวอย่าง n ในการทดสอบ μ กรณีไม่ทราบ σ^2 และ n<30

H_o : $\mu = \mu_0$

H_1 : $\mu \neq \mu_0$ $(\mu_1 = \mu_0 + d)$

$$\text{Non-central t} = \frac{\bar{x} - (\mu o + d)}{S/\sqrt{n}}$$

หา n โดยใช้ตารางภาคผนวกที่ A8 เมื่อทราบ

$$\Delta = \frac{|\mu - \mu o|}{\sigma} \quad \text{และ } \alpha, \beta$$

ใช้ S แทน σ

$$\mu = \mu o + d$$

$$\therefore \quad \Delta = \frac{|d|}{S}$$

กรณี Two Tails ใช้ $\frac{\alpha}{2}, \beta$ กรณี One Tail ใช้ α, β

5.15 การหาขนาดตัวอย่าง n ในการทดสอบ μ_1-μ_2 กรณีไม่ทราบ σ_1^2, σ_2^2 และ $n_1, n_2 < 30$

Ho : $\mu_1 - \mu_2 = 0$

H_1 : $\mu_1 \neq \mu_2 \neq 0$ $(\mu_1 - \mu_0 = d)$

ถ้า $\sigma_1^2 = \sigma_2^2$

ให้ $$S^2 = \sigma_1^2 = \sigma_2^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

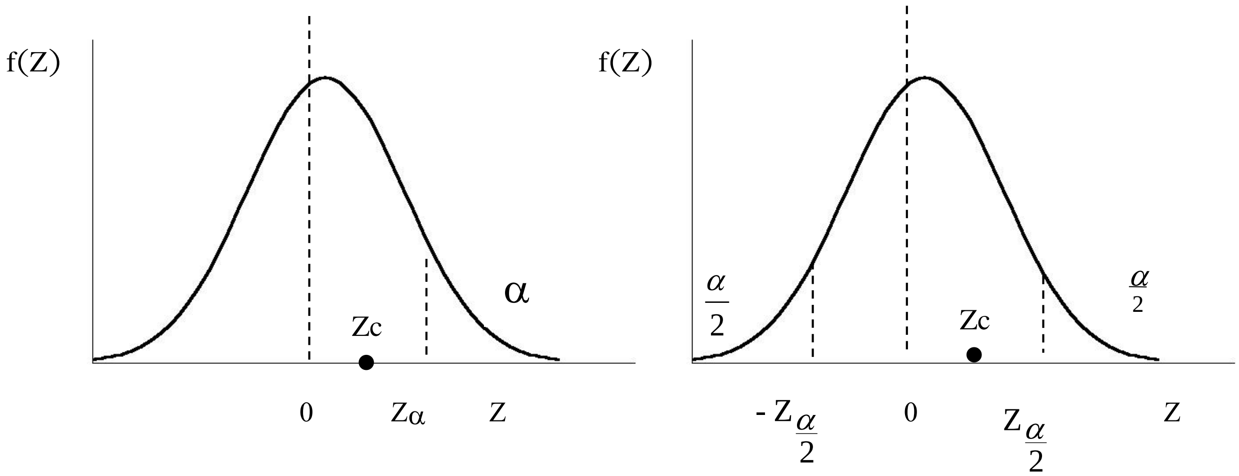
$$\Delta = \frac{|d|}{S}$$

เหมือนข้อ 5.14 แต่ใช้ตารางต่างกัน

5.16 P-value (Probability Value)

หมายถึง โอกาสความน่าจะเป็นน้อยที่สุดที่สามารถปฏิเสธ Ho

ถ้า p-value $< \alpha$ จะ Reject Ho ดังรูปที่ 5.7



รูปที่ 5.7 การทดสอบสมมติฐานโดยใช้ p-value

$$z_c = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$ มีค่าทั้ง + และ –

p-value $= P(z > |z_c|)$ ....(One Tail)

p-value $= 2 * P(z > |z_c|)$ .... (Two Tails)

$= P(z \le -z_c$ และ $z \ge z_c)$

5.17 การใช้ χ^2 ในการทดสอบแบบจำลอง (Chi-Square Tests of Models)

ในหัวข้อที่แล้วได้กล่าวถึงการใช้ χ^2 ทดสอบความแปรปรวนของประชากร 1 ชุด และได้นิยาม χ^2 ไว้ดังนี้

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{S^2}{\sigma^2/\upsilon}$$

υ = k-1 (กรณีรู้พารามิเตอร์ของประชากร)

υ = k-1-m (กรณีไม่รู้ค่าพารามิเตอร์ของประชากร แต่ต้องประมาณค่าพารามิเตอร์ จำนวน m ตัว จากตัวอย่าง)

ในบทนี้จะได้กล่าวถึงการใช้ χ^2 ในการทดสอบกรณีต่าง ๆ ดังต่อไปนี้

- ทดสอบความเหมาะสมของการแจกแจงหรือแบบจำลอง (Test of Goodness of Fit)
- ทดสอบความเป็นอิสระของเหตุการณ์หรือ Random Variable (Test for Independence)
- ทดสอบสัดส่วน, p (Test for Proportion)
- ทดสอบความเหมือนกันของประชากร (Test for Homogeneity)

5.17.1 การ Derive หาค่า χ^2

ถ้า X มีการแจกแจงแบบ Binomial มี b(x; n, p) และมีคุณสมบัติดังตาราง

Outcome	No. of Success	Prob. of Success
E_i (Success) E_2 (Failure)	x_1 $x_2 = n-x_1$	p_1 $p_2 = 1-p_1$
Total	= n	= 1

ถ้า n มาก , X $\approx$ Approximate Normal ซึ่งสามารถแปลงเป็น z ได้ดังสมการ

$$z = \frac{x_1 - np_1}{\sqrt{np_1(1-p_1)}}$$

$$z^2 = \frac{(x_1 - np_1)^2}{np_1(1-p_1)}$$

$$= \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_1 - np_1)^2}{n(1-p_1)}$$

ถ้า $x_1 - np_1 = (n-x_2)-n(1-p_2)$

$$= \quad -x_2+np_2 = \quad -(x_2-np_2)$$

$$(x_1-np_1)^2 = \quad (x_2-np_2)^2$$

$$z^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2}$$

$$z^2 = \chi^2 = \sum_{i=1}^{k-2} \frac{(x_i - np_i)^2}{np_i} \qquad \text{.....(5.1)}$$

มี υ = k-m-1

ถ้า X_i มีการแจกแจงแบบ Multinomial มีผลลัพธ์ k แบบ มี $f(x_1, x_2, \ldots x_k ; p_1, p_2, \ldots p_k, n)$ และมีคุณสมบัติดังตาราง

Outcomes	No. of Success	Prob
E_1	x_1	p_1
E_2	x_2	p_2
.	.	.
.	.	.
.	.	.
E_k	x_k	p_k
	n	1.00

ในทำนองเดียวกับสมการ 4.1

$$Q_{k-1} = z^2 = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i} \qquad \text{.....(5.2)}$$

ถ้า n มีค่ามาก จนทำให้ $np_i > 5$ สำหรับทุกค่า i

$Q_{k-1} \approx \chi^2$ ซึ่งมี $\upsilon = k-1$

ถ้าให้ x_i = observed frequency จากการทดลอง = O_i

และ np_i = expected frequency ตามทฤษฎี = e_i

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \qquad \text{.....(5.3)}$$

ใช้สำหรับการทดสอบว่าผลการทดลอง (O_i) ต่างจากค่าตามทฤษฎี (e_i) อย่างมีนัยสำคัญทางสถิติหรือไม่

ถ้าโมเดลตามทฤษฎีมีพารามิเตอร์ m ตัว ซึ่งไม่ทราบค่า และต้องประมาณค่าจากสถิติ χ^2 จะมี Degree of Freedom (υ)

$$\upsilon = k-m-1 \qquad \text{……………………………..(5.4)}$$

5.17.2 การทดสอบความเหมาะสมของการแจกแจง (Test of Goodness of Fit)

คือการทดสอบว่าข้อมูลที่รวบรวมมามีการแจกแจงตามที่คาดไว้หรือไม่ เช่น มีการแจกแจงแบบปกติจริงหรือไม่ หรือผลการทดลองเป็นไปตามโมเดล (ทฤษฎี) ที่ตั้งไว้หรือไม่

มีขั้นตอนดังนี้

(1) ตั้งสมมติฐาน

Ho : ข้อมูลมีการแจกแจงตามที่คาดไว้ เช่น $X \sim N\ (x; \mu,\sigma)$
(ผลการทดลองเป็นไปตามโมเดล (ทฤษฎี) ที่ตั้งไว้)

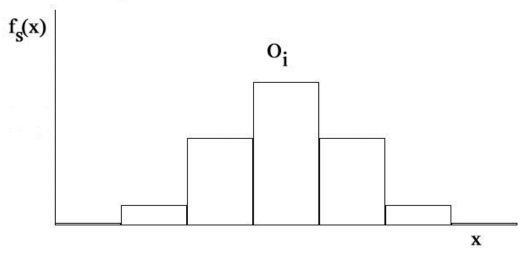
H_1 : ข้อมูลไม่ได้มีการแจกแจงตามที่คาดไว้
(ผลการทดลองไม่ได้เป็นไปตามโมเดล (ทฤษฎี) ที่ตั้งไว้)

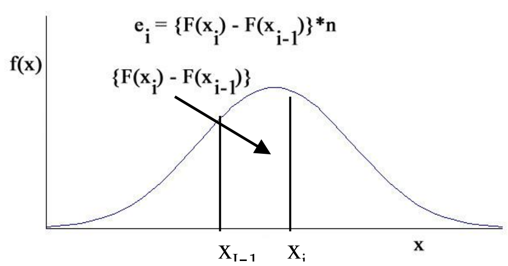
(2) นำผลที่ได้จากการทดลองมาแจกแจงความถี่ [O_i]

(3) คำนวณความถี่ที่คาดว่าจะเป็นตามทฤษฎี (ตาม Ho) [e_i]

ถ้า $e_i < 5$ ให้รวม Class ใกล้กันเข้าด้วยกันเพื่อให้ $e_i \geq 5$

(4) เลือก α เพื่อเปรียบเทียบความแตกต่างของข้อมูล (O_i VS. e_i) ดูรูปที่ 5.8





รูปที่ 5.8 การเปรียบเทียบ O_i กับ e_i

คำนวณ $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{O_i^2}{e_i} - N$

(5) ทดสอบด้วย χ^2 (One tail ด้านบนอย่างเดียว)

ถ้า $\chi^2 > \chi^2_{\alpha,\, k-m-1}$ ให้ Reject H_o

ถ้า $\chi^2 < \chi^2_{\alpha,\, k-m-1}$ ให้ Accept H_o

ตัวอย่างที่ 5.16 สุ่มตัวอย่างเลข 0-9 แบบใส่คืนจำนวน 51 ครั้งได้ผลดังนี้

5	8	3	1	9	4	6	7	9	2
6	3	0	8	7	5	1	3	6	2
1	9	5	4	8	0	3	7	1	4
6	0	4	3	8	2	1	3	9	8
5	6	1	8	7	0	3	5	2	5
2									

จงทดสอบว่า การสุ่มตัวอย่างถูกต้องตามทฤษฎี หรือไม่

ถ้าการสุ่มตัวอย่างถูกต้องตามทฤษฎี

P (ได้เลขเดิม) = $\frac{1}{10}$

P (ได้เลขติดต่อ) = $\frac{2}{10}$

P (ได้เลขอื่น) = $\frac{7}{10}$

H_o : การสุ่มเลขเป็นไปตามทฤษฎี

H_1 : การสุ่มไม่เป็นไปตามทฤษฎี

α = 0.05

υ = k-m-1 = 3-0-1 = 2

χ^2 = $\chi^2_{0.05,2}$ = 5.991

ผลลัพธ์	ความถี่จากการทดลอง (ครั้ง) O_i	ความถี่ตามทฤษฎี (ครั้ง) e_i
ได้เลขเดิม	0	$50 \times \frac{1}{10} = 5$
ได้เลขติดกัน	8	$50 \times \frac{2}{10} = 10$
ได้เลขอื่น	42	$50 \times \frac{7}{10} = 35$
รวม	50	50

ตรวจสอบว่า $e_i > 5$ ทุกค่า

$$\chi^2 = \frac{(0-5)^2}{5} + \frac{(8-10)^2}{10} + \frac{(42-35)^2}{35}$$

$$= 5+0.1+\frac{(7)^2}{35} = 8$$

$$\chi^2 > \chi^2_{0.05,\,2}$$

Reject H_o หรือสรุปว่าเลขสุ่มทั้ง 51 ตัวไม่เป็นไปตามทฤษฎีที่ระดับนัยสำคัญ 0.05

ตัวอย่างที่ 5.17 จงทดสอบว่าอัตราการไหลของของไหลชนิดหนึ่งมีการแจกแจงแบบปกติ ที่ระดับความเชื่อมั่น 95% จากข้อมูลการวัดอัตราการไหล (x) 80 ครั้ง ซึ่งมีการแจกแจงดังตาราง

อัตราการไหล x (lps)						
ชั้น	1	2	3	4	5	6
ขอบเขตชั้น	7.5-7.8	7.8-8.1	8.1-8.4	8.4-8.7	8.7-9.0	9.0-9.3
ความถี่ (ครั้ง)	5	21	35	15	3	1

การคำนวณค่า $\bar{x}$ และ s

ชั้น (i)	ขอบเขตชั้น (lps)	จุดกึ่งกลางชั้น x_i	ความถี่ f	$f.x_i$	$f.x_i^2$
1	7.5-7.8	7.65	5	35.25	292.61
2	7.8-8.1	7.95	21	166.95	1327.25
3	8.1-8.4	8.25	35	288.25	2382.19
4	8.4-8.7	8.55	15	128.25	1096.54
5	8.7-9.0	8.85	3	26.25	234.97
6	9.0-9.3	8.16	1	9.15	83.72
		$\sum$	80	657.60	5417.28

$$\bar{x} = \frac{\sum_{i=1}^{6} f\ x_i}{\sum_{i=1}^{6} f} = \frac{657.6}{80} = 8.22$$

$$s^2 = \frac{n\sum f.x_i^2 - (\sum f.x_i)^2}{n(n-1)}$$

$$= \frac{80 \times 5417.25 - (657.6)^2}{80 \times 79}$$

s^2 = 0.1495

s = 0.3866

H_o = Q มีการแจกแจงปกติ

H_1 = Q ไม่มีการแจกแจงปกติ

α = 0.05

การคำนวณค่า e_i

ชั้น(i)	ขอบเขตชั้น	z_i	$e_i=n*P(z_{i-1}< z < z_i)$
1	7.5-7.8	< -1.0864	80(0.1386) = 11.09
2	7.8-8.1	-0.3103	80(0.3771-0.1386) = 19.08
3	8.1-8.4	0.4656	80(0.6792-0.3771) = 24.17
4	8.4-8.7	1.2416	80(0.8927-0.6792) = 17.08
5	8.7-9.0	2.0176	80(0.9782-0.8927) = 6.87
6	9.0-9.3	>2.0176	80(1.-0.9782) = 1.74

การคำนวณค่า χ^2

ชั้น(i)	ขอบเขตชั้น	O_i	e_i	$(O_i-e_I)^2/ei$
1	7.5 - 7.8	5	11.09	3.34
2	7.8 - 8.1	21	19.08	0.19
3	8.1 - 8.4	35	24.17	4.85
4	8.4 - 8.7	15	17.08	0.25
$\left.\begin{matrix}5\\6\end{matrix}\right\}5$	8.7 - 9.0 9.0 - 9.3	$\left.\begin{matrix}3\\1\end{matrix}\right\}4$	$\left.\begin{matrix}6.84\\1.74\end{matrix}\right\}8.58$	$\}2.44$
$\sum$		80	80	11.07

$m = 2$ (ใช้ $\bar{x} \sim \mu$ และใช้ $S \sim \sigma$)

$\chi^2_{\alpha,\ k-m-1} = \chi^2_{0.05,\ 5-2-1} = 5.99$

$\chi^2\ (11.07) > \chi^2_{0.05,\ 2}\ (5.99)$

Reject H_o หรือสรุปได้ว่า X มีการแจกแจงแบบปกติที่ $\alpha = 0.05$

ตัวอย่างที่ 5.18 จงทดสอบว่าจำนวนสินค้าชำรุดในแต่ละวันมีการแจกแจงแบบ Poisson หรือไม่ จากผลการตรวจสอบสินค้าชำรุด 100 วันดังตาราง

จำนวนสินค้าชำรุด x (ชิ้น/วัน)	0	1	2	3	4	5	6
จำนวนวัน f	28	43	10	5	5	2	7

กำหนดให้ $\alpha = 0.05$

H_o = จำนวนผลิตภัณฑ์ชำรุดในแต่ละวันมีการแจกแจงแบบ Poisson

H_1 = จำนวนผลิตภัณฑ์ชำรุดในแต่ละวันไม่เป็น Poisson

Poisson distribution

$$P(x;\mu) = \frac{e^{-\mu}\mu^{x}}{X!} \quad \text{เมื่อ } x = 0, 1, 2, \ldots\ldots$$

$$P(x \le r) = P(r;\mu) = \sum_{x=0}^{r} P(x;\mu)$$

คำนวณค่า $\bar{x}$

$$\bar{x} = \frac{\sum_{i=0}^{6} fx_i}{\sum_{i=0}^{6} f} = \frac{150}{100} = 1.5$$

$$P(r; 1.5) = \sum_{x=0}^{r} \frac{e^{-1.5}\, 1.5^{x}}{x!}$$

จำนวนผลิตภัณฑ์ชำรุด /วัน (x)	O_i	$P(x; 1.5) = \frac{e^{-1.5}\, 1.5^{x}}{x!}$	e_i	$\frac{(O_i - e_i)^2}{e_i}$
0	28	0.2231	22.31	1.451
1	43	0.3347	33.47	2.714
2	10	0.2510	25.1	9.084
3	5	0.1255	12.55	4.542
4	5	0.0471	4.71	
5	2 } 14	0.0141	1.41 } 6.47	8.764
6	7	0.0035	0.35	
$\sum$	100	0.999	99.9	26.55

$\chi^2 = 26.55$

$\chi^2_{\alpha,\, \upsilon = k-m-1} = \chi^2_{0.05,\, 5-1-1} = \chi^2_{0.05,\, 3}$

$= 7.81$

$\chi^2\ (26.55) > \chi^2_{0.05,\, 3}\ (7.81)$

Reject H_o หรือสรุปได้ว่าจำนวนสินค้าชำรุดในแต่ละวันไม่ได้มีการแจกแจงแบบ Poisson ที่ $\alpha = 0.05$

ตัวอย่างที่ 5.19 ผลการสำรวจครอบครัวที่มีบุตร 5 คน จำนวน 320 ครอบครัวได้ผลดังตาราง

จำนวนบุตร เพศชาย/หญิง	จำนวนครอบครัว
5/0	18
4/1	56
3/2	110
2/3	88
1/4	40
0/5	8
$\sum$	320

จงทดสอบว่าข้อมูลมีการแจกแจงแบบ Binomial และโอกาสการเกิดบุตรเพศชายและหญิงเท่ากันที่ระดับนัยสำคัญเท่ากับ 0.05

การแจกแจงแบบ Binomial

$$b(x\,;\,n,p) = \binom{n}{x} p^{x} q^{1-x} \quad \text{เมื่อ} \quad x = 0, 1, 2, \ldots, n$$

$$= \frac{n!}{x!\,(n-x)\,!}\; p^{x} q^{1-x}$$

H_o = จำนวนบุตรชาย/หญิง มีการแจกแจงแบบ Binomial และโอกาสการเกิดบุตรเพศชายและหญิงเท่ากับ (p= 0.5)

H_1 = ไม่ใช่ Binomial และ $p \neq q \neq 0.5$

จำนวนบุตรชาย/หญิง x/(n-x)	จำนวนครอบครัว (O_i)	$e_i = \binom{n}{x} 0.5^{x}\, 0.5^{n-x}$ xn	$\frac{(O_i - e_i)^2}{e_i}$
5/0	18	$(0.5)^5(0.5)^0$x320 = 10	6.4
4/1	56	$5(0.5)^4(0.5)$x320 = 50	0.72
3/2	110	$10(0.5)^3(0.5)^2$x320 = 100	1
2/3	88	$10(0.5)^2(0.5)^3$x320 = 100	1.44
1/4	40	$5(0.5)^1(0.5)^4$x320 = 50	2
0/5	8	$(0.5)^0(0.5)^5$x320 = 10	0.4
$\sum$	320	320	11.96

χ^2 = 11.96

$\chi^2_{\alpha,\ k\text{-}m\text{-}1}$ = $\chi^2_{0.05,\ 6\text{-}0\text{-}1}$; (m=0 เนื่องจากรู้ n และ p)

= $\chi^2_{0.05,\ 5}$ = 11.07

χ^2 (11.96) > $\chi^2_{0.05,5}$ (11.07)

Reject H_o หรือสรุปว่าจำนวนบุตรชายหญิงมีการแจกแจงแบบ Binomial และ $p \neq q \neq 0.5$ ที่ α = 0.05

ตัวอย่างที่ 5.20 ในการหยิบไพ่ 3 ใบ จากสำหรับไพ่ 52 ใบ หยิบเสร็จจดไพ่โพธิ์ดำแล้วคืนที่เดิม ทำการทดลองซ้ำกัน 64 ครั้งได้ผลดังตาราง

จำนวนไพ่โพธิ์ดำที่หยิบได้ (ใบ)	0	1	2	3
จำนวนครั้ง	16	36	12	0

จงทดสอบว่าข้อมูลมีการแจกแจงแบบ Binomial ที่ α = 0.05

p = 0.25

H_o = จำนวนไพ่โพธิ์ดำมีการแจกแจงแบบ Binomial

H_1 = จำนวนไพ่โพธิ์ดำไม่มีการแจกแจงแบบ Binomial

จำนวนโพธิ์ดำ, x (ใบ)	O_i	b (x;3,0.25) = $\binom{n}{x} p^x q^{1-x}$	e_i = b (x;3,0.25)	$\frac{(O_i - e_i)^2}{e_i}$
0	16	0.4219	27	4.481
1	36	0.4219	27	3
2	12 }	0.1406	9 }	}
3	0 }	0.0156	1 } 10	} 0.4

$\chi^2 = 7.8815$

$\chi^2_{\alpha,\ k\text{-}m\text{-}1} = \chi^2_{0.05,\ 3\text{-}0\text{-}1}$

$= \chi^2_{0.05,\ 2} = 5.991$

$\chi^2\ (7.8815) = \chi^2_{0.05,\ 2}\ (5.991)$

Reject H_o หรือสรุปว่าจำนวนโพธิ์ดำมีการแจกแจงแบบ Binomial ที่ $\alpha = 0.05$

5.17.3 การทดสอบความเป็นอิสระ (Test for Independence)

ถ้า A และ B เป็นอิสระต่อกัน ; $P(A \cap B) = P(A).\ P(B)$

ถ้า A และ B ไม่เป็นอิสระต่อกัน ; $P(A \cap B) = P(B).P(A/B) = P(A).P(B/A)$

การทดสอบความเป็นอิสระ คือการทดสอบว่าคุณลักษณะ/เหตุการณ์ 2 จำพวก เป็นอิสระต่อกัน หรือมีความสัมพันธ์กัน เช่น

* **การเลือกคณะ**ในการสอบเข้ามหาวิทยาลัย [B] ขึ้นอยู่กับ**เพศ** [A] หรือไม่

- **ถ้าขึ้นอยู่กับเพศ** แสดงว่า การเลือกคณะกับเพศ**ไม่เป็นอิสระ**ต่อกัน (มี Condition)
- **ถ้าไม่ขึ้นกับเพศ** แสดงว่า การเลือกคณะกับเพศ**เป็นอิสระ**ต่อกัน (ไม่มี Condition)

* **มะเร็งปอด** [B] ขึ้นอยู่กับ**การสูบบุหรี่** [A] หรือไม่

- **ถ้าขึ้น** แสดงว่า มะเร็งปอดกับการสูบบุหรี่**ไม่เป็นอิสระ**ต่อกัน (มี Condition)
- **ถ้าไม่ขึ้น** แสดงว่า มะเร็งปอดกับการสูบบุหรี่**เป็นอิสระ**ต่อกัน (ไม่มี Condition)

ขั้นตอนการทดสอบ

(1) ตั้งสมมติฐาน

Ho : คุณลักษณะ/เหตุการณ์ทั้งคู่**เป็นอิสระ**ต่อกัน

H_1 : ไม่เป็นอิสระ (มีความสัมพันธ์)

(2) นำผลการสำรวจ/ทดลองจำนวน N มาแจกแจงความถี่ให้อยู่ในรูปของตารางจร (Contingency Table) [O_{ij}] ตารางแสดงความถี่ของการเกิดเหตุการณ์ $A \cap B$ จากตัวอย่าง

(3) คำนวณความถี่ที่คาดว่าจะเป็นตามทฤษฎี ถ้าข้อมูลเป็นอิสระต่อกันตาม Ho [e_{ij}]

$e_{ij} \nless 1$ และ $e_{ij} < 5$ ต้องไม่น้อยกว่า 20% ของ rc

(4) เลือก α เพื่อเปรียบเทียบว่า 2 และ 3 ต่างกันอย่างมีนัยสำคัญหรือไม่

(5) ทดสอบด้วย $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-e_{ij})^2}{e_{ij}}$ ถ้า $\chi^2 > \chi^2_{\alpha,\upsilon}$ Reject Ho

ตารางจร (Contingency Table) หรือ Joint Frequency ของคุณลักษณะ/เหตุการณ์จากตัวอย่าง 2 จำพวก

c = จำนวนคุณลักษณะ/เหตุการณ์ B ที่ต้องการทดสอบว่าขึ้นอยู่กับ A หรือไม่ (Dependent/Independent)

r = จำนวนลักษณะ/เหตุการณ์ (A) ที่ต้องการทดสอบว่ามีผลต่อ B หรือไม่ (Independent)

	B_1	B_2	.	B_j	.	B_C	Marginal Frequency
A_1	O_{11}	O_{12}		O_{1j}		O_{1c}	R_1
A_2	O_{21}	O_{22}		O_{2j}		O_{2c}	R_2
.	.	.		.		.	.
.	.	.		.		.	.
A_j	O_{i1}	O_{i2}		O_{ij}		O_{ic}	R_i
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
Ar	O_{r1}	O_{r2}		O_{rj}		O_{rc}	R_r
Marginal Frequency	C_1	C_2	.	C_j	.	C_C	N

Ho : A และ B เป็นอิสระต่อกัน (B ไม่ขึ้นกับ A)

H_1 : A และ B ไม่อิสระต่อกัน (B ขึ้นกับ A)

ตามสมมติฐาน H_1

$P(Ai \cap Bj) = P(Bj)\,P(Ai/Bj) = P(Ai)\,P(Bj/Ai)$

ตามสมมติฐาน Ho

$$P(Ai \cap Bj) = P(Ai)\,P(Bj)$$

$$= \frac{Ri}{N} \cdot \frac{Cj}{N}$$

$$e_{ij} = P(Ai \cap Bj)\,N$$

$$= \frac{Ri}{N} \cdot \frac{Cj}{N} \cdot N = \frac{Ri.\,Cj}{N}$$

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{O^2{}_{ij}}{e_{ij}} - N$$

$e_{ij} \nless 1$

$e_{ij} < 5$ ต้องไม่เกิน 20% ของ rc

$k = rc$

$m = (c-1)+(r-1) = c+r-2$

$\upsilon = k-m-1$

$= rc-c-r+2-1$

$= rc-c-r+1$

$= (r-1)(c-1)$

วิธีการคำนวณ χ^2 โดยตารางจร

	B_1	B_2	.	B_j	.	B_C	MF
A_1	O_{11} / e_{11}	O_{12} / e_{12}		O_{1j} / e_{1j}		O_{1c} / e_{1c}	R_1
A_2	O_{21} / e_{21}						R_2
.							.
A_i	O_{i1} / e_{i1}			O_{ij} / e_{ij}		O_{ic} / e_{ic}	R_i
.							.
A_r	O_{r1} / e_{r1}			O_{rj} / e_{rj}		O_{rc} / e_{rc}	R_r
MF	C_1	C_2	.	C_j	.	C_C	N

ตัวอย่างที่ 5.21 ในการทดสอบว่าการเลือกคณะในการสอบเข้ามหาวิทยาลัยขึ้นอยู่กับเพศหรือไม่ ที่ระดับระดับนัยสำคัญ (α) เท่ากับ 0.01 ได้ทำการการสุ่มนิสิต 400 คน ได้ผลดังตาราง

เพศ (A)	คณะ (B)					รวม
	บริหาร	วิศวฯ	อักษร	พยาบาล	เภสัช	
ชาย	21	16	145	2	6	190
หญิง	14	4	175	13	4	210
รวม	35	20	320	15	10	400

H_o : การเลือกคณะไม่ขึ้นกับเพศ (independent)

H_1 : การเลือกคณะขึ้นกับเพศ

α = 0.01

คำนวณ e_{ij} ถ้า H_o เป็นจริง

$$e_{ij} = \frac{R_i , C_j}{N}$$

เพศ	คณะ					รวม
	บริหาร	วิศวฯ	อักษร	พยาบาล	เภสัช	
ชาย	$190 \times \frac{35}{400}$ = 16.625	$190 \times \frac{20}{400}$ = 9.50	$190 \times \frac{320}{400}$ = 152	$190 \times \frac{15}{400}$ = 7.125	$190 \times \frac{10}{400}$ = 4.75	190
หญิง	$210 \times \frac{35}{400}$ = 18.375	$210 \times \frac{20}{400}$ = 10.50	$210 \times \frac{320}{400}$ = 168	$210 \times \frac{15}{400}$ = 7.875	$210 \times \frac{10}{400}$ = 5.250	210
รวม	35	20	320	15	10	N=400

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{5} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(21-16.625)^2}{16.625} + \frac{(16 - 9.50)^2}{9.50} + ... + \frac{(4 - 5.25)^2}{5.25}$$

$$= 1.15 + 4.45 + 0.32 + 3.69 + 0.33 + 1.04 + 4.02 + 0.29 + 3.34 + 0.3$$

= 18.93

$\chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{0.01,4} = 13.28$

$\chi^2(18.93) > \chi^2_{0.01,4}(13.28)$

Reject H_o หรือสามารถสรุปได้ว่าการเลือกคณะขึ้นอยู่กับเพศที่ระดับนัยสำคัญ 0.01 (ไม่ Independent)

ตัวอย่างที่ 5.22 จงทดสอบว่าความร้ายแรงของอุบัติเหตุทางรถยนต์มีความสัมพันธ์กับการใช้อุปกรณ์นิรภัยหรือไม่ที่ระดับนัยสำคัญเท่ากับ 0.01 ได้ทำการสุ่มตัวอย่างจากทะเบียนประวัติอุบัติเหตุ 1,000 ราย ได้ผลดังตาราง

ความร้ายแรงของอุบัติเหตุ (B)	อุปกรณ์นิรภัย (A)			รวม
	เข็มขัด 1	เข็มขัดและอื่น ๆ 2	ไม่ใช้ 3	
1. ไม่บาดเจ็บ	75	60	65	200
2. บาดเจ็บเล็กน้อย	160	115	175	450
3. อาการหนัก	100	65	135	300
4. ตาย	15	10	25	50
รวม	350	250	400	1,000

H_o : ความร้ายแรงของอุบัติเหตุไม่มีความสัมพันธ์กับการใช้อุปกรณ์นิรภัย (Independent)

H_1 : ความร้ายแรงของอุบัติเหตุมีสัมพันธ์กับการใช้อุปกรณ์นิรภัย

α = 0.01

$$e_{ij} = \frac{R_i, C_j}{N}$$

	เข็มขัด	เข็มขัดและอื่น ๆ	ไม่ใช้	รวม
ไม่บาดเจ็บ	$\frac{200 \times 350}{1000}$ = 70	$\frac{200 \times 250}{1000}$ = 50	$\frac{200 \times 400}{1000}$ = 80	200
บาดเจ็บเล็กน้อย	157	112.5	180	450
อาการหนัก	105	75	120	300

ตาย	17.5	12.5	20	50
รวม	350	250	400	1,000

$$\chi^2 = \sum_{i=1}^{4}\sum_{j=1}^{3}\frac{(O_{ij}-e_{ij})^2}{e_{ij}}$$

$$= \frac{(75-70)^2}{70}+\frac{(60-50)^2}{50}+.......+\frac{(25-20)^2}{20}$$

$$= 10.96$$

$$\chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{0.1,(4-1)(3-1)}$$

$$= \chi^2_{0.1,6} = 10.6$$

$$\chi^2\ (10.96) > \chi^2_{0.1,6}\ (10.6)$$

Reject H_o หรือสรุปได้ว่าความร้ายแรงของอุบัติเหตุมีความสัมพันธ์กับการใช้อุปกรณ์นิรภัย (ไม่ Independent)

5.17.4 การทดสอบความเหมือนกันของประชากร (Test for Homogeneity)

ในการทดสอบความเหมือนกันของประชากรจะพิจารณาว่าสำหรับคณะใด ความเห็นของชาย/หญิงในการเลือกคณะเหมือนกัน (Homogeneity) หรือต่างกัน (Heterogeneity)

จากตารางจรในข้อ 5.20 (Test for Independence) เปลี่ยนสัญญลักษณ์จาก O_{ij} เป็น n_{ij}

	B_1	B_2		B_j		B_C	MF*
A_1	n_{11}	n_{12}		n_{1j}		n_{1c}	$n_{1.}$ (กำหนดล่วงหน้า)
A_2	n_{21}	n_{22}		n_{2j}		n_{2c}	$n_{2.}$ (กำหนดล่วงหน้า)
.	.	.		.		.	.
A_i	n_{i1}	n_{i2}		n_{ij}		n_{ic}	$n_{i.}$ (กำหนดล่วงหน้า)
.	.	.		.		.	.
A_r	n_{r1}	n_{r2}		n_{rj}		n_{rc}	$n_{r.}$ (กำหนดล่วงหน้า)
MF	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.c}$	n (กำหนดล่วงหน้า)

ให้ $p_{ij} = \dfrac{n_{ij}}{n_{i.}}$ = proportion (เมื่อ i = 1, 2,.........,r และ j = 1, 2,..........., c)

หมายเหตุ $n_{i.}$ (เมื่อ i = 1, 2, ..., r) เป็นค่าที่กำหนดไว้ล่วงหน้าก่อนสุ่มตัวอย่างซึ่งต่างจากการ Test for Independence ค่า $n_{i.}$ ไม่ได้กำหนดไว้ล่วงหน้า

<u>ตารางจรข้างบนจะเปลี่ยนเป็น</u> (เปลี่ยน n_{ij} เป็น p_{ij})

	B_1	B_2	.	B_j	.	B_C	รวม
A_1	$p_{11} = \frac{n_{11}}{n_{1.}}$	p_{12}		p_{1j}		p_{1c}	1
A_2	$p_{21} = \frac{n_{21}}{n_{2.}}$	p_{22}		p_{2j}		p_{2c}	1
A_i	$p_{i1} = \frac{n_{i1}}{n_{i.}}$	p_{i2}		p_{ij}		p_{ic}	1
A_r	$p_{r1} = \frac{n_{r1}}{n_{r.}}$	p_{r2}		p_{rj}		p_{rc}	1
	$p_{.1} = \frac{n_{.1}}{n}$	$p_{.2} = \frac{n_{.2}}{n}$	.	$p_{.j} = \frac{n_{.j}}{n}$	.	$p_{.c} = \frac{n_{.c}}{n}$	1.00

ขั้นตอนการทดสอบ

1. ตั้งสมมติฐาน [Ho : ที่ B_j ใด ๆ proportion ของ B_j/A_i (เมื่อ i = 1, 2, ..., r) เท่ากัน และเท่ากับ proportion ของ B_j ดังนั้น $P(B_j/A_i) = P(B_j)$ หรือ B_j independent จาก A_i]

Ho : สำหรับ B_j ใด $p_{1j} = p_{2j} = ... = p_{ij} = ... = p_{rj} = p_{.j}$
เมื่อ j = 1,2,...., c (Homogeneity)

H_1 : สำหรับ B_j ใด proportion ไม่เท่ากันทุกตัว หรืออย่างน้อยคู่หนึ่งไม่เท่ากัน เมื่อ j = 1,2,....., c (Heterogeneity)

2. นำผลการสำรวจ/ทดลองมาทำตารางจร (n_{ij})

3. คำนวณหา $e_{ij} = n_i \times p_{ij}$ ตาม Ho ($p_{ij} = p_{.j}$; $e_{ij} = n_{i.} \times p_{.j} = \frac{n_{i.} \times n_{.j}}{n}$)

4. กำหนด α

5. ทดสอบด้วย $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$

ถ้า $\chi^2 > \chi^2_{\alpha, (r-1)(c-1)}$ Reject Ho

ความจริงวิธีการทดสอบ Homogeneity เหมือนการทดสอบ Independence ทุกประการ เพียงแต่ต้องการทดสอบว่าสำหรับ B_j ที่กำหนดให้ใด ๆ proportion ของ Bj/Ai (i=1, 2,....., r) เท่ากันหรือไม่ ถ้าผลสรุปว่าเท่ากัน จะสามารถคำนวณหา proportion ของ B_j/A_i ได้จาก proportion ของ B_j [$P(B_j/A_i) = P(B_j)$ แสดงว่า B_j และ Ai Independent]

$$p_{ij} = p_{.j} = \frac{n_{.j}}{n} \quad \text{เมื่อ } i = 1, 2, \ldots\ldots, r \text{ และ } j = 1, 2, \ldots\ldots, c$$

ตัวอย่างที่ 5.23 จงทดสอบว่าสัดส่วนต้นไม้มีระดับคลอโรฟินเท่ากัน ถึงแม้ว่าระดับ SO_2 แตกต่างกัน จากผลการสุ่มตัวอย่างดังตาราง

ระดับของ SO_2 (A)	ระดับคลอโรฟิน (B)			รวม
	สูง	กลาง	ต่ำ	
สูง	3 / 5	4 / 8.33	13 / 6.67	20
กลาง	5 / 5	10 / 8.33	5 / 6.67	20
ต่ำ	7 / 5	11 / 8.33	2 / 6.67	20
	15	25	20	60

$$e_{ij} = \frac{n1.\,x\,n.1}{n} = \frac{20\,x\,15}{60} = 5$$

$$= e_{21} = e_{31}$$

$$e_{12} = \frac{n1.\,x\,n.2}{n} = \frac{20\,x\,25}{60} = 8.33$$

$$= e_{22} = e_{32}$$

$$e_{13} = \frac{n1.\,x\,n.3}{n} = \frac{20\,x\,20}{60} = 6.67$$

$$= e_{23} = e_{33}$$

H_o : $p_{1j} = p_{2j} = p_{3j}$ $(j=1, 2, 3)$

H_1 : $p_{1j} \neq p_{2j} \neq p_{3j}$

α = 0.01

$\chi^2_{\alpha,(r-1)(c-1)}$ = $\chi^2_{0.01,2x2}$

$$= \chi^2_{0.01,4}$$
$$= 13.3$$

$$\chi^2 = \frac{(3-5)^2}{5} + \frac{(4-8.33)^2}{8.33} + ... + \frac{(11-8.33)^2}{8.33} + \frac{(2-6.67)^2}{6.67}$$
$$= 14.74$$
$$> \chi^2_{0.01,4}\,(13.3)$$

Reject H_o หรือ $p_{1j} \neq p_{2j} \neq p_{3j}$

แสดงว่าจากการสุ่มตัวอย่าง ระดับ SO_2 ในบรรยากาศรอบ ๆ ต้นพืชนั้นไม่เหมือนกัน (ไม่ Homogeneous) จึงมีผลทำให้ระดับคลอไรฟินในต้นไม้มีค่าต่างกัน หรือ SO_2 ในบรรยากาศรอบ ๆ ต้นพืชมีผลต่อคลอโรฟินในต้นไม้

5.17.5 การทดสอบสัดส่วน (Test for Proportion)

การทดสอบสัดส่วนเหมือนกับ Test for Homogeneity แต่ใช้ทดสอบ k Independent Binomial Samples หรือในกรณีที่ B มีผลลัพธ์ 2 แบบ (Binomial) (หรือ r=2) เท่านั้น เช่นสัดส่วนชาย/หญิงที่เลือกคณะวิศวกรรมศาสตร์ (หรือไม่วิศวกรรมศาสตร์) เท่ากัน

Ho : สัดส่วนของชาย/หญิง ในการเลือกคณะวิศวกรรมศาสตร์เหมือนกัน

($p_{11} = p_{21}$ หรือ $p_{12} = p_{22}$)

H_1 : สัดส่วนของชาย/หญิงในการเลือกคณะวิศวกรรมศาสตร์ไม่เท่ากัน

($P_{11} \neq P_{21}$ หรือ $p_{12} \neq p_{22}$)

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

ถ้าผลการทดสอบยอมรับ Ho จะสรุปว่าสัดส่วนชาย/หญิงเท่ากันในการเลือกคณะวิศวกรรมศาสตร์หรือไม่เลือกวิศวกรรมศาสตร์

ตัวอย่างที่ 5.24 จงทดสอบว่าสัดส่วน (Proportion) ของผู้มีโลหิตเปลี่ยนแปลงของประชากร 2 กลุ่มเท่ากัน จากข้อมูลผลการตรวจสอบโลหิตของประชากรที่อาศัยอยู่ใกล้แหล่งที่มีสารกัมมันตภาพรังสี 300 คน ไกลแหล่ง 320 คน ดังตาราง

	โลหิตเปลี่ยนแปลง (D)	โลหิตไม่เปลี่ยนแปลง (D')	รวม
ประชากรใกล้แหล่ง (E)	52	248	300
ประชากรไกลแหล่ง (E')	48	272	320
รวม	100	520	620

H_o : $p_{11} = p_{21}$

H_1 : $p_{11} \neq p_{21}$

$\alpha = 0.25$

$$\chi^2 = \chi^2_{\alpha, (r-1)(c-1)}$$

$$= \chi^2_{0.25, (2-1)(2-1)}$$

$$= \chi^2_{0.25,1} = 1.32$$

	D	D'	รวม
E	52 / 48.39	248 / 251.61	300
E'	48 / 51.62	272 / 268.39	320
	100	520	620

$$e_{11} = \frac{n1.\,x\,n.1}{n} = \frac{300 \times 100}{620}$$

$$= 48.38$$

$$e_{21} = \frac{n2.\,x\,n.1}{n} = \frac{320 \times 100}{620} = 51.62$$

$$e_{12} = \frac{n1.\,x\,n.2}{n}$$

$$= \frac{300 \times 520}{620} = 251.61$$

$$e_{22} = \frac{n2.\,x\,n.2}{n}$$

$$= \frac{320 \times 520}{620} = 268.38$$

$$\chi^2 = \sum_{i=1}^{r=2}\sum_{j=1}^{c=2}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}$$

$$= \frac{(52-48.39)^2}{48.39}+\frac{(248-251.61)^2}{251.61}$$

$$= \frac{(48-51.62)^2}{51.62}+\frac{(272-268.39)^2}{268.39}$$

$$= 0.62$$

$$< \chi^2_{0.25,\,1}\ (1.32)$$

Accept H_o : $p_{12} = p_{21}$ หรือสรุปได้ว่าสัดส่วนผู้มีโลหิตเปลี่ยนแปลงของประชากรที่อยู่ใกล้แหล่งที่มีสารกัมมันตภาพรังสีและไกลแหล่งเท่ากันที่ระดับนัยสำคัญ 0.25

จากตารางขนาด 2x2

a	b	a+b
c	d	c+d
a+c	b+d	a+b+c+d

สูตรอย่างง่าย $\chi^2 = \dfrac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$

$N = a+b+c+d$

ถ้า $5 < e_{ij} < 10$ ต้องมีการปรับค่า (Adjustment) โดยวิธี Yate's Correction

$$\chi^2_{corrected} = \frac{N[|ad-bc|-\frac{1}{2}N]^2}{(a+b)(a+c)(b+d)(c+d)}$$

จากตัวอย่างที่ 9.9

$$\chi^2 = \frac{620\,(52 \times 272 - 48 \times 248)^2}{300 \times 100 \times 320 \times 520} = 0.62$$

เนื่องจาก $e_{ij} > 10$ จึงไม่ต้องมี Correction

5.18 แบบฝึกหัด

1. ถุงใบหนึ่งมีลูกแก้วสีแดงและสีเขียว จงทดสอบสมมติฐานว่ามีลูกแก้วสองสีนี้ในอัตราส่วนเท่ากัน โดยหยิบลูกแก้วจากถุงทีละลูก 30 ครั้ง แบบหยิบแล้วใส่คืนก่อนหยิบครั้งต่อไป และตั้งเกณฑ์ว่าจะยอมรับสมมติฐานถ้าหยิบได้ลูกแก้วสีเขียวตั้งแต่ 12 ถึง 18 ลูก
   ก. จงหาความน่าจะเป็นที่จะไม่ยอมรับสมมติฐาน เมื่อความจริงสมมติฐานนั้นถูกต้อง
   ข. จงหาอำนาจการทดสอบ เมื่อความจริงในถุงนั้นมีลูกแก้วสีแดงต่อสีเขียวในอัตราส่วน 2:3

2. มหาวิทยาลัยแห่งหนึ่งอ้างว่า มีนักศึกษาซึ่งมีสภาพวิทยาภัณฑ์เพียง 10% เพื่อทดสอบคำกล่าวดังกล่าว จึงตั้งเกณฑ์ไว้ว่า ถ้าจำนวนนักศึกษามีสภาพวิทยาทัณฑ์มากกว่า 3 คน จากจำนวน 15 คนที่สุ่มมา จะปฏิเสธคำกล่าวอ้าง
   ก. จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1
   ข. จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 2 เมื่อความจริงแล้วนักศึกษามีสภาพวิทยาทัณฑ์เท่ากับ 20%
   ค. จงหาอำนาจการทดสอบเมื่อนักศึกษามีสภาพทัณฑ์ 30% เปรียบเทียบกับอำนาจการทดสอบเมื่อนักศึกษามีสภาพวิทยาทัณฑ์ 20% (หาค่าความน่าจะเป็นที่แท้จริง)
   ง. เมื่อสุ่มนักศึกษามา 600 คน พบว่า 64 คน มีสภาพวิทยาทัณฑ์ จะสรุปผลการทดสอบนี้อย่างไร ด้วยระดับนัยสำคัญ 0.05

3. บริษัทผู้ผลิตรถยนต์ได้ทำการคิดค้นยางรถยนต์ชนิดใหม่ชนิดหนึ่ง โดยอ้างว่า ระยะการใช้งานเฉลี่ยมีค่าอย่างน้อยเท่ากับ 40,000 กิโลเมตร เพื่อทดสอบคำกล่าวอ้างของบริษัทผู้ผลิตนี้ วิศวกรผู้หนึ่งจึงทำการสุ่มตัวอย่างยางชนิดใหม่นี้มา 10 เส้น และทำการทดสอบได้ข้อมูลดังต่อไปนี้

ยางเส้นที่	1	2	3	4	5	6	7	8	9	10
ระยะการใช้งาน (พันกิโลเมตร)	36.1	40.2	33.8	38.5	42.0	35.8	37.0	41.0	36.8	37.2

   จงทดสอบคำกล่าวอ้างของบริษัทผู้ผลิต โดยใช้ระดับนัยสำคัญ 0.05

4. ในการทดสอบสมมติฐานว่าค่าเฉลี่ยอายุการใช้งานของหลอดไฟฟ้าที่ผลิตจากโรงงานแห่งหนึ่งมีค่ามากกว่า 800 ชั่วโมงหรือไม่ ที่ระดับนัยสำคัญ 0.1 ถ้าต้องการอำนาจการทดสอบเป็น 95% และทราบว่าค่าความแปรปรวนของอายุการใช้งานของหลอดไฟนี้เป็น 1,600 ชั่วโมง^2 สมมติให้การแจกแจง

ของอายุการใช้งานของหลอดไฟเป็นแบบปกติ และค่าเฉลี่ยที่แท้จริงของอายุการใช้งานของหลอดไฟนี้มีค่าเป็น 835 ชั่วโมง จงหาขนาดตัวอย่างที่เหมาะสม

5. วิศวกรผู้หนึ่งต้องการตรวจปริมาณสารพิษ PCB ที่เจือปนอยู่ในน้ำ จึงได้ทำการจับปลา 10 ตัว โดยสุ่มแบบไม่เจาะจงจากทะเลสาบ A และทำการตรวจวัดระดับความเข้มข้นของสาร PCB ด้วยเครื่องมือหนึ่ง ได้ข้อมูลดังนี้ (หน่วย ppm ; part per millon)

   ทะเลสาบ A : 11.5, 10.8, 11.6, 9.4, 12.4, 11.4, 12.2, 11, 10.6, 10.8

   นอกจากนี้เขาได้ทำการสุ่มจับปลาอีก 8 ตัวจากทะเลสาบ B ซึ่งอยู่ใกล้เคียงกัน และทำการตรวจวัดระดับความเข้มข้นของสารพิษ PCB ด้วยเครื่องมืออีกชนิดหนึ่งซึ่งแตกต่างจากที่ใช้กับการตรวจที่ทะเลสาบ A ได้ข้อมูลดังต่อไปนี้ (หน่วย ppm)

   ทะเลสาบ B : 11.8, 12.6, 12.2, 11.7, 12.1, 10.4, 12.6

   จากข้อมูลที่ผ่านมาพบว่า เครื่องมือที่ใช้ในการตรวจวัดที่ทะเลสาบ A มีความแปรปรวนเท่ากับ 0.09 ขณะที่เครื่องมือที่ใช้ที่ทะเลสาบ B มีความแปรปรวนเท่ากับ 0.16 จะสรุปได้หรือไม่ว่าทะเลสาบทั้งสองแห่งนี้มีปริมาณสารพิษ PCB ในระดับความเข้มข้นที่แตกต่างกัน ใช้ระดับนัยสำคัญ 0.05

6. มีการโต้เถียงกันในเรื่องความสามารถในการต้านทานกระแสไฟฟ้าของลวด A ว่ามีความต้านทานสูงกว่าลวด B วิศวกรจึงทำการทดสอบลวดทั้งสองชนิดโดยสุ่มตัวอย่างชนิดละ 6 เส้น ได้ข้อมูลดังนี้

แรงต้านทาน (โอห์ม)	
ลวด A	ลวด B
0.140	0.135
0.138	0.140
0.143	0.136
0.142	0.142
0.44	0.138
0.137	0.140

ท่านจะทำการสรุปจากข้อมูลนี้ได้อย่างไรเพื่อจะได้ยุติข้อโต้เถียงนี้ ใช้ระดับนัยสำคัญ 0.1

7. หญิงมีครรภ์ 10 คน ได้ถูกสุ่มมาเพื่อทำการทดสอบโดยฉีดยาชนิดหนึ่งที่มีผลต่อการบีบตัวของมดลูกขณะคลอด โดยที่แพทย์จะทำการเก็บรวบรวมข้อมูลความดันโลหิตก่อนและหลังการฉีดยา และบันทึกไว้ในตารางต่อไปนี้

คนไข้ที่	ก่อนฉีดยา	หลังฉีดยา	คนไข้ที่	ก่อนฉีดยา	หลังฉีดยา
1	134	140	6	140	138
2	122	130	7	118	124
3	132	135	8	127	126
4	130	126	9	125	132
5	128	134	10	142	144

จงทดสอบว่า ฤทธิ์ของยาชนิดนี้มีผลต่อค่าความดันโลหิตของสตรีในขณะตั้งครรภ์หรือไม่ ใช้ระดับนัยสำคัญ 0.05

8. สมมติว่าสุ่มตัวอย่างขนาด $n_1 = n_2 = n$ จากประชากรปกติ 2 ชุด ซึ่งมีค่าความแปรปรวน $\sigma^2_1 = 101$ และ $\sigma^2_2 = 94.7$ เพื่อทดสอบสมมติฐานหลัก $H_o : \mu_1 - \mu_2 = 3$ แย้งกับสมมติฐานหลัก $H_1 : \mu_1 - \mu_2 > 3$ กำหนดให้ความน่าจะเป็นของการเกิดความผิดพลาดประเภทที่ 1 เท่ากับ 0.05 และอำนาจการทดสอบ 99% จงหาค่าขนาดตัวอย่างที่เหมาะสม

9. กระบวนการเคมีหนึ่งสามารถใช้ตัวกระตุ้นปฏิกิริยาของกระบวนการได้ 2 ชนิด ในการทดสอบว่าความแปรปรวนของผลที่ได้จากกระบวนการนี้เมื่อใช้ตัวกระตุ้นปฏิกิริยาทั้งสองไม่แตกต่างกันนัก เคมีผู้หนึ่งจึงได้ทำการสุ่มตัวอย่างขนาด $n_1 = 10$ จากกระบวนการที่ใช้ตัวกระตุ้นปฏิกิริยาชนิดที่ 1 และทำการสุ่มตัวอย่างขนาด $n_2 = 12$ จากกระบวนการที่ใช้ตัวกระตุ้นปฏิกิริยาชนิดที่ 2 จากตัวอย่างทั้งสองชุดนี้พบว่า $S^2_1 = 0.14$ หรือ $S^2_2 = 0.28$ อยากพบว่าที่ระดับนัยสำคัญ 0.05 ท่านจะสรุปได้หรือไม่ว่า ความแปรปรวนของผลที่ได้จากกระบวนการนี้ เมื่อใช้ตัวกระตุ้นปฏิกิริยาชนิดที่ 1 มีค่าต่ำกว่าความแปรปรวนของผลที่ได้จากกระบวนนี้ เมื่อใช้ตัวกระตุ้นปฏิกิริยาชนิดที่ 2

10. บริษัทผู้ผลิต computer chip แห่งหนึ่งอ้างว่า จำนวน chip ที่บกพร่องที่ผลิตจากโรงงานมีจำนวนน้อยกว่า 2% ของจำนวน chip ที่ผลิตทั้งหมด บริษัทผลิตเครื่องอิเล็กทรอนิกส์แห่งหนึ่งต้องการที่จะซื้อ chip เป็นจำนวนมาก จึงต้องการตรวจสอบคำกล่าวอ้างของบริษัทผู้ผลิต computer chip รายนี้ โดยการสั่งซื้อ chip มาจำนวนหนึ่งและสุ่มตัวอย่าง chip 300 ชิ้น เพื่อทำการทดสอบ ถ้าพบ chip ที่บกพร่อง

จำนวน 10 ชิ้น บริษัทผู้ผลิตเครื่องอิเล็กทรอนิกส์ควรจะสั่งซื้อ chip จากบริษัทผู้ผลิต computer chip รายนี้หรือไม่ ทดสอบที่ระดับนัยสำคัญ 0.01

11. ทฤษฎีของเมนเดลชี้ให้เห็นว่าลักษณะและสีของเมล็ดถั่วต่าง ๆ สามารถแบ่งเป็นกลุ่มใหญ่ ได้ 4 กลุ่ม คือ

กลุ่มที่ 1 : "เมล็ดกลมและมีสีเหลือง"

กลุ่มที่ 2 : "เมล็ดกลมและมีสีเขียว"

กลุ่มที่ 3 : "เมล็ดหักงอและมีสีเหลือง"

และกลุ่มที่ 4 : "เมล็ดหักงอและมีสีเขียว"

ในอัตราส่วน 9:3:3:1 สุ่มตัวอย่างเมล็ดถั่วมา 556 เมล็ด ทำการจำแนกเมล็ดพันธุ์ตามกลุ่มต่าง ๆ ได้ผลดังตาราง

กลุ่มที่	จำนวนที่สังเกตได้ (เมล็ด)
1	315
2	108
3	101
4	32

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่าข้อมูลที่ได้จากตัวอย่างดังกล่าวสอดคล้องกับทฤษฎีของเมนเดล

12 ในการสำรวจความคิดเห็นทางการเมือง เพื่อตรวจสอบว่า ผู้มีสิทธิ์ออกเสียงเลือกตั้งที่มีอายุต่ำกว่า 25 ปี มีความคิดเห็นที่แตกต่างไปจากผู้มีสิทธิ์ออกเสียงเลือกตั้งที่มีอายุตั้งแต่ 25 ปีขึ้นไป หรือไม่ จากการสัมภาษณ์ผู้มีสิทธิ์ออกเสียง 1,500 คน ที่มีอายุตั้งแต่ 25 ปีขึ้นไป และ 1,000 คน ที่มีอายุต่ำกว่า 25 ปี ได้ผลความคิดเห็นดังตาราง

อายุ (ปี)	ไม่เห็นด้วย	ไม่แน่ใจ	เห็นด้วย	ทั้งหมด
< 25	400	100	500	1,000
ตั้งแต่ 25 ขึ้นไป	600	400	500	1,500
ทั้งหมด	1,000	500	1,000	2,500

จงทดสอบสมมติฐานว่า ไม่มีเหตุผลเพียงพอที่จะกล่าวได้ว่าผู้มีสิทธิ์ออกเสียงทั้งสองกลุ่มมีความคิดเห็นที่แตกต่างกันทางการเมือง ที่ระดับนัยสำคัญ 0.01

13. ในการสำรวจลักษณะทางพันธุกรรมเกี่ยวกับโรคตาบอดสีว่าเป็นไปตามทฤษฎีทางพันธุกรรมดังตารางที่ 1 หรือไม่ ($q = 1-p$) นักวิจัยได้ทำการทดสอบคน 1,000 คน ได้ผลดังตารางที่ 2

ตารางที่ 1

เพศ / ลักษณะ	ชาย	หญิง
ปกติ	$p/2$	$p^2/2+pq$
ตาบอดสี	$q/2$	$q^2/2$

ตารางที่ 2

เพศ / ลักษณะ	ชาย	หญิง
ปกติ	442	514
ตาบอดสี	38	6

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่าลักษณะทางพันธุกรรมเกี่ยวกับโรคตาบอดสีจากตัวอย่างชุดนี้ว่าเป็นไปตามทฤษฎีพันธุกรรมหรือไม่ (ทดสอบโดยใช้ $p = 0.8$)

14. ในการทดสอบผลของการใช้เซรุ่มป้องกันโรคหวัดชนิดหนึ่ง ได้ทำการทดลองใช้เซรุ่มนี้กับคน 500 คน บันทึกผลเป็นเวลา 1 ปี เพื่อเปรียบเทียบกับกลุ่มตัวอย่างคนที่ไม่ได้รับเซรุ่มชนิดนี้500 คน ได้ผลดังตารางต่อไปนี้

	ไม่เป็นหวัดเลย	เป็นหวัด 1 ครั้ง	เป็นหวัด > 1 ครั้ง
ได้รับเซรุ่ม	252	145	103
ไม่ได้รับเซรุ่ม	224	136	140

จงทดสอบที่ระดับนัยสำคัญ 0.05 ว่า การใช้เซรุ่มไม่มีผลต่อการป้องกันหวัดชนิดนี้

15. นักวิจัยผู้หนึ่งได้ทำการจำแนกนักศึกษา 1,725 คนออกเป็นกลุ่ม ๆ ตามความสามารถในการเรียน และฐานะทางครอบครัว โดยได้ผลดังตารางต่อไปนี้

	สมองทึบ	ฉลาด	ฉลาดมาก
ฐานะดีมาก	81	322	233
ฐานะดี	141	457	153
ฐานะยากจน	127	163	78

จงทดสอบที่ระดับนัยสำคัญ 0.01 ว่าความสามารถในการเรียนไม่ขึ้นกับฐานะทางครอบครัว

16. โรงงานแห่งหนึ่งใช้เครื่องจักรในกระบวนการผลิตสินค้าชนิดหนึ่งพร้อมกัน 4 เครื่อง และในแต่ละวันมีการทำงานทั้งหมด 3 ผลัด พบว่าจำนวนครั้งที่เครื่องจักรชำรุด (Breakdown) ในช่วง 6 เดือนที่ผ่านมามีข้อมูลดังตาราง

เครื่องจักร / ผลัดที่	A	B	C	D	ผลรวม/ผลัด
1	10	12	6	7	35
2	10	24	9	10	53
3	13	20	7	10	50
ผลรวม/เครื่องจักร	33	56	22	27	138

จงทดสอบที่ระดับนัยสำคัญ 0.05 ว่าการชำรุดของเครื่องจักรและผลัดของงานที่มีการชำรุดเกิดขึ้นนั้นเป็นอิสระต่อกัน พร้อมทั้งระบุค่า p-value

17. จากข้อมูลในอดีตของเครื่องจักรเครื่องหนึ่ง ชี้ให้เห็นว่า สินค้าแต่ละชิ้นที่ทำการผลิตจากเครื่องจักรนี้สามารถจำแนกเป็นเกรดต่าง ๆ ดังนี้

เกรด A (คุณภาพดีมาก) ด้วยความน่าจะเป็น 0.40
เกรด B (คุณภาพดี) ด้วยความน่าจะเป็น 0.30
เกรด C (คุณภาพปานกลาง) ด้วยความน่าจะเป็น 0.20
เกรด D (คุณภาพต่ำ) ด้วยความน่าจะเป็น 0.10

ถ้าโรงงานทำการซื้อเครื่องจักรใหม่เครื่องหนึ่ง ซึ่งนำไปใช้ในการผลิตสินค้าชนิดเดียวกับเครื่องจักรเดิม ถ้าเครื่องจักรเครื่องใหม่นี้ทำการผลิตสินค้าชนิดนั้น 500 ชิ้น ได้ข้อมูลดังตาราง

เกรด	A	B	C	D
จำนวนชิ้น	234	117	81	68

อยากทราบว่า เครื่องจักรใหม่มีคุณภาพในการผลิตแตกต่างกันจากเครื่องจักรเดิมหรือไม่ ที่ระดับนัยสำคัญ 0.01

18. จากการตรวจสอบอายุการใช้งานของท่อเครื่องดูดฝุ่นยี่ห้อหนึ่ง โดยการสุ่มตัวอย่างท่อเครื่องดูดฝุ่นนี้ 100 อัน มาทำการทดสอบได้ข้อมูลอายุการใช้งานดังตาราง

อายุการใช้งาน (ชั่วโมง)	จำนวนท่อ
< 30	41
ระหว่าง 30 ถึง 60	34
ระหว่าง 60 ถึง 90	13
> 30	15

จงทดสอบที่ระดับนัยสำคัญ 0.05 ว่าข้อมูลชุดนี้สนับสนุนคำกล่าวที่ว่า อายุการใช้งานของท่อเครื่องดูดฝุ่นยี่ห้อนี้มีการแจกแจงแบบเอ็กซ์โปเนนเซียล ด้วยค่าเฉลี่ยอายุการใช้งานเท่ากับ 50 ชั่วโมง พร้อมระบุค่า p-value

19. ในการทดลองเพื่อศึกษาความสัมพันธ์ระหว่างความเครียดกับการสูบบุหรี่ได้ทำการเก็บรวบรวมข้อมูลดังตาราง

	ไม่สูบบุหรี่	สูบปานกลาง	สูบหนัก
ความเครียดสูง	20	38	28
ไม่มีความเครียด	50	27	18

จากข้อมูลดังกล่าว จงทดสอบสมมติฐานว่า ความเครียดในแต่ละบุคคลไม่มีผลต่อปริมาณการสูบบุหรี่ของแต่ละบุคคลนั้น ทดสอบที่ระดับนัยสำคัญ 0.05

20. ในการศึกษาพบว่า จำนวนครั้งของการเกิดไฟดับในแต่ละวันในเมือง ๆ หนึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปัวซองด้วยค่าเฉลี่ย 4.2 จงทดสอบว่าข้อมูลไฟฟ้าดับที่เก็บในช่วงเวลา 150 วันดังตาราง มีความสอดคล้องกับผลการศึกษาดังกล่าวข้างต้นหรือไม่ ทดสอบที่ระดับนัยสำคัญ 0.05

จำนวนครั้งที่ไฟดับ	0	1	2	3	4	5	6	7	8	9	10	11
จำนวนวันที่เกิดเหตุการณ์นั้น	0	5	22	23	32	22	19	13	6	4	4	0

บทที่ 6
การถดถอยเชิงเส้นและสหสัมพันธ์ (Linear Regression and Correlation)

6.1 แนวความคิดในการวิเคราะห์การถดถอย (Regression Analysis)

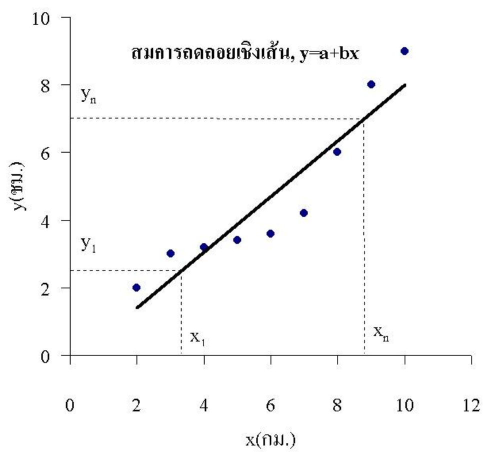
(1) ต้องการศึกษาว่าการขับรถจาก มหาวิทยาลัยเกษตรศาสตร์ (บางเขน) ไปตามสถานที่ต่าง ๆ ใน กทม. ใช้เวลามากน้อยเท่าใด โดยการทดลองจับเวลาและระยะทางในการขับรถ

ให้ x = ระยะทาง (กม)

Y = เวลา (ชม)

(ผลการทดลอง)

x (กม)	Y (ชม)
x_1 x_2 x_3 . . . x_n	y_1 y_2 y_3 . . . y_n
ตัวแปรอิสระ/ต้น Independent Variable	ตัวแปรตาม Dependent Variable
Not Random Variable	Random Variable
ปกติ กำหนด/ควบคุม ได้	กำหนด/ควบคุมไม่ได้ (ที่ x ใด ๆ มี y ได้หลาย ค่า ซึ่งไม่สามารถกำหนดหรือควบคุมได้)



* **ตามปกติ** ระยะทางมากกว่าจะใช้เวลาในการขับรถมากกว่าตามสมการ

กรณีนี้ สามารถกำหนดกลับกันได้ เช่น ให้ y เป็นตัวแปรอิสระ (Not Random Variable) และ x เป็นตัวแปรตาม (Random Variable) ได้

แต่ $x \neq -\frac{a}{b} + \frac{1}{b}y$

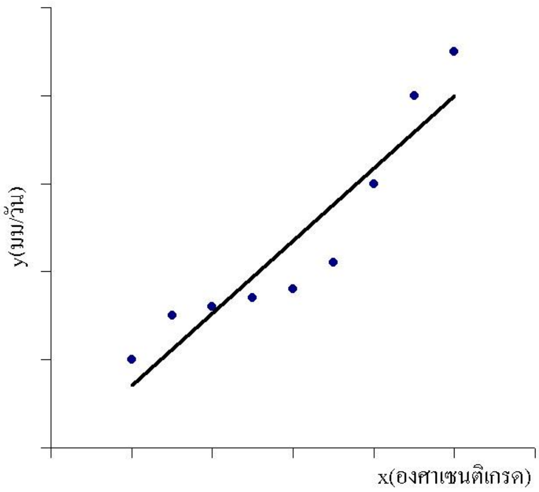
(2) ต้องการศึกษาว่าอุณหภูมิของอากาศ มีผลในเชิงตัวเลขต่อปริมาณการใช้น้ำของพืชอย่างไร

x = อุณหภูมิอากาศ$^\circ$C (วัดได้/กำหนดหรือควบคุมไม่ได้)

Y = ปริมาณการใช้น้ำของพืช (มม/วัน)

(ผลการทดลอง)

x($^\circ$c)	Y (มม./วัน)
x_1	y_1
x_2	y_2
.	.
.	.
.	.
x_n	y_n



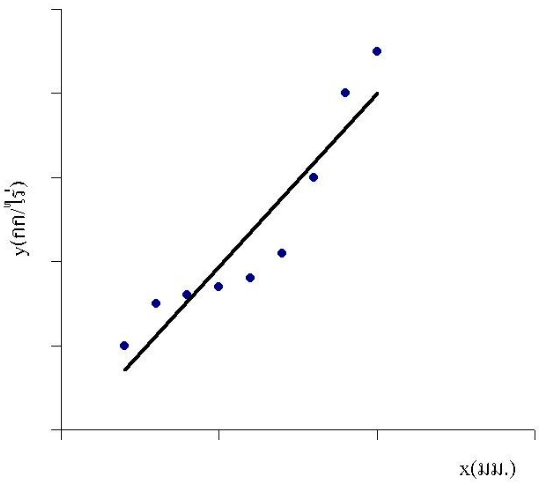
(3) ต้องการศึกษาว่าปริมาณน้ำชลประทาน มีผลในเชิงตัวเลขต่อผลผลิตข้าวโพดอย่างไร

x = ปริมาณน้ำชลประทานที่ให้ (มม.) (วัด/กำหนด/ควบคุมได้)

Y = ผลผลิตข้าวโพด (กก./ไร่)

(ผลการทดลอง)

x(มม.)	Y (กก./ไร่)
x_1	y_1
x_2	y_2
.	.
.	.
.	.
x_n	y_n



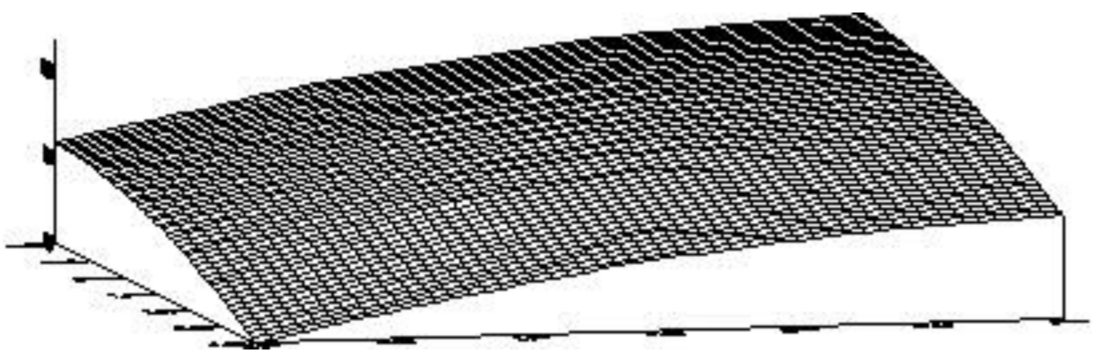
(4) ต้องการศึกษาว่าปริมาณน้ำชลประทาน และปุ๋ย มีผลในเชิงตัวเลขต่อผลผลิตข้าวโพดอย่างไร

x = ปริมาณน้ำชลประทานที่ให้ (มม.)
y = ปริมาณปุ๋ยเคมีที่ให้ (กก./ไร่) กำหนด/ควบคุมได้
Z = ผลผลิตข้าวโพด (กก./ไร่)

(ผลการทดลอง)

x (มม)	y (กก./ไร่)	Z (กก./ไร่)
x_1	y_1	z_{11}
	y_2	z_{12}
	.	.
	y_n	z_{1n}
x_2	y_1	z_{21}
	y_2	z_{22}
	.	.
	y_n	z_{2n}
.	.	.
x_m	y_1	z_{m1}
	y_2	z_{m2}
	.	.
	y_n	z_{mn}

Response Surface
(not linear)



6.2 Simple และ Multiple Regression

(1) Simple (แบบง่าย)

Y VS. x เช่นกรณี (1), (2) และ (3) ในหัวข้อ 6.2

(2) Multiple (พหุคูณ)

Z VS. x, y

หรือ Y VS. x_1, x_2, ..., x_n

(3) รูปแบบความสัมพันธ์ระหว่าง y และ x

$y = a+bx$ (Linear/เชิงเส้น/เส้นตรง)

$y = a+\frac{b}{x}$ (Hypergeometric/Reciprocal)

$y = a+bx+cx^2$ (Polynomial)

$y = e^{a+bx}$ (Exponential) ($y=ae^{bx}$)

$y = a\,x^b$ (Power)

$y = \frac{x}{ax+b}$ (Hyperbolic)

(a, b, c = พารามิเตอร์ของสมการ)

6.3 การถดถอยเชิงเส้นแบบง่าย (Simple Linear Regression)

$Y = a+bx$

ให้ Y = Dependent Random Variable

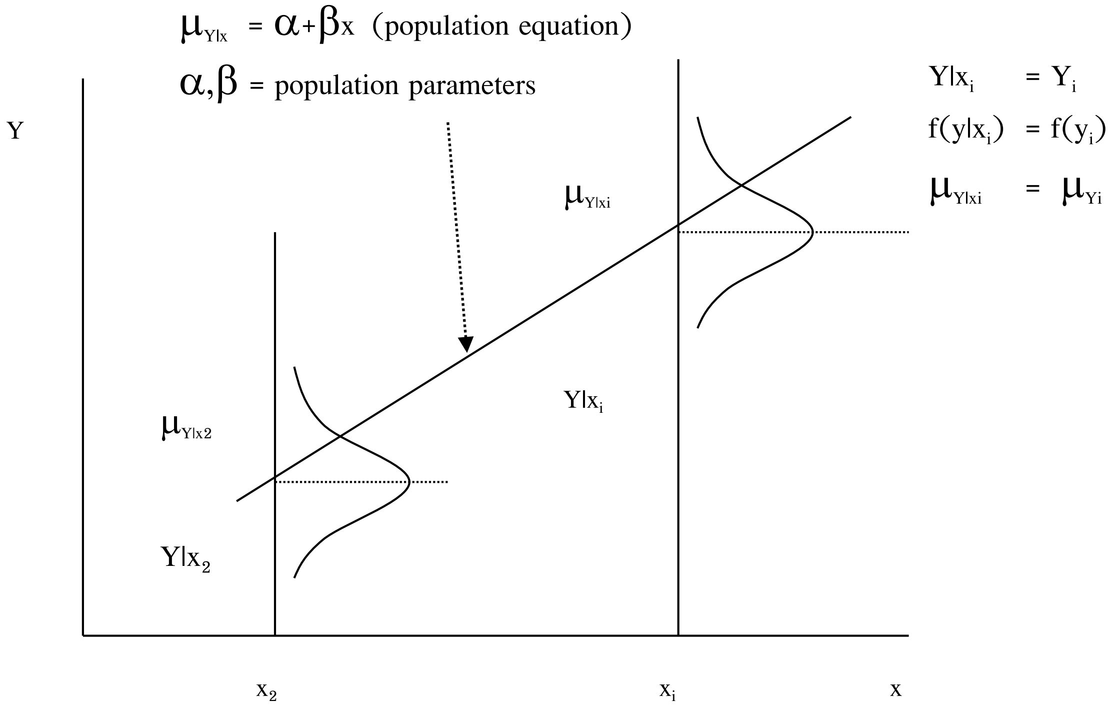
x = Independent Variable

ผลการทดลองหรือสุ่มตัวอย่างได้ (x_i, y_i) เมื่อ $i = 1, 2, ..., n$

$y_i \in Y_i$(ค่าของ Random Variable เมื่อ $x = x_i$)

หรือ $y_i \in Y|xi$ ซึ่งมี Mean = $\mu_{Y|xi}$ และ Variance = $\sigma^2_{Y|xi}$

ความสัมพันธ์ระหว่าง x และ y สามารถแสดงไว้ดังรูปที่ 6.1



รูปที่ 6.1 ความสัมพันธ์เชิงเส้นระหว่าง x และ y

ที่ x_i ใด ๆ

$Y|x_i \quad = \quad \mu_{Y|xi} + Ei$

$\quad\quad\;\; = \quad \alpha + \beta x_i + Ei$

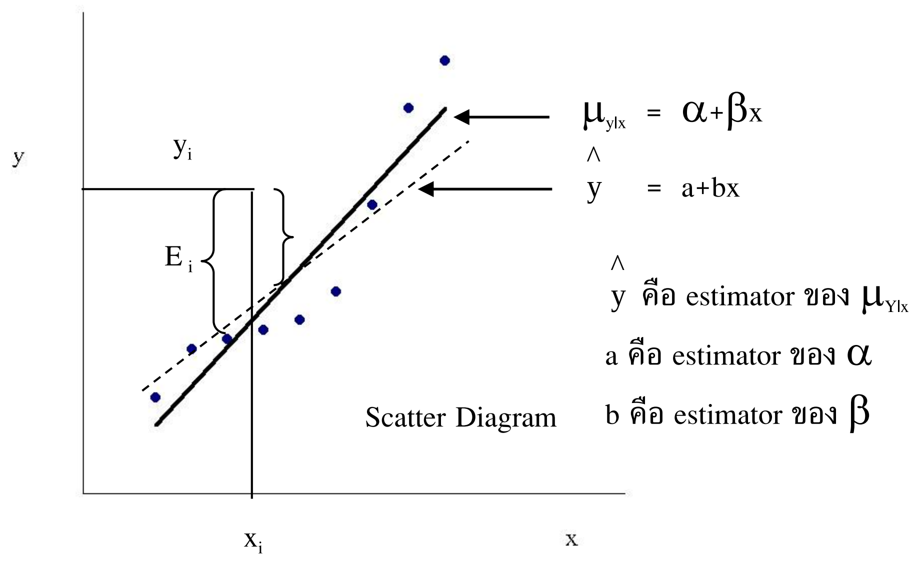
$Ei \quad = \quad$ Random หรือ Model Error ซึ่งมี $\mu_{Ei} = 0$ และ $\sigma_{Ei} = \sigma$

จาก Sample $\{(x_i, y_i) ; i = 1,2,......, n\}$ สามารถพล๊อต Scatter Diagram ได้ดังรูปที่ 10.2

$\hat{y} \quad = \quad a+bx \quad (\hat{} = \text{ estimate}) \sim (\mu_{Y|x} = \alpha+\beta x)$

ที่ x_i ใด ๆ $\quad y_i \quad = \quad \hat{y}_i + e_i \; = \; a+bx_i+e_i$

$e_i \quad = \quad$ residual หรือ error in fitted model



รูปที่ 6.2 Scatter Diagram

6.3.1 Least Square Estimate

วิธีการหาค่าสัมประสิทธิ์ a และ b ของสมการถดถอยเชิงเส้นแบบง่ายซึ่งทำให้

SSE = sum of square of error (sum of square of residual)

= มีค่าน้อยที่สุด

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

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

$$\frac{\partial SSE}{\partial a} = 0 \rightarrow a\Sigma_1 + b\Sigma x_i = \Sigma y_i \quad \text{..................................(6.1)}$$

$$\frac{\partial SSE}{\partial b} = 0 \rightarrow a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_i y_i \quad \text{..................................(6.2)}$$

สมการปกติ (Normal Equation)

จาก (6.1) $$a = \frac{\sum y_i}{n} - b\frac{\sum x_i}{n} = \bar{y} - b\bar{x} \quad \text{........(6.3)}$$

แทนค่า a ลงใน (6.2) $$b = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$= \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad \text{...(6.4)}$$

ตัวอย่างที่ 6.1 จงหาความสัมพันธ์ระหว่าง y (การระเหยของสีเป็น % โดยน้ำหนัก) และ x (% ความชื้นสัมพันธ์) จากข้อมูลดังตาราง

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	35.3	11.0	9	70.7	7.8	17	59.3	10.1
2	29.7	11.1	10	57.5	9.1	18	70.0	8.1
3	30.8	12.5	11	46.4	8.2	19	70.0	6.8
4	58.8	8.4	12	28.9	12.2	20	74.4	8.9
5	61.4	9.3	13	28.1	11.9	21	72.1	7.7
6	71.3	8.7	14	39.1	9.6	22	58.1	8.5
7	74.4	6.4	15	46.8	10.9	23	44.6	8.9
8	76.7	8.5	16	48.5	9.6	24	33.4	10.4
					n =	25	28.6	11.1

$\sum x_i = 1{,}314.90 \qquad \sum y_i = 235.70$

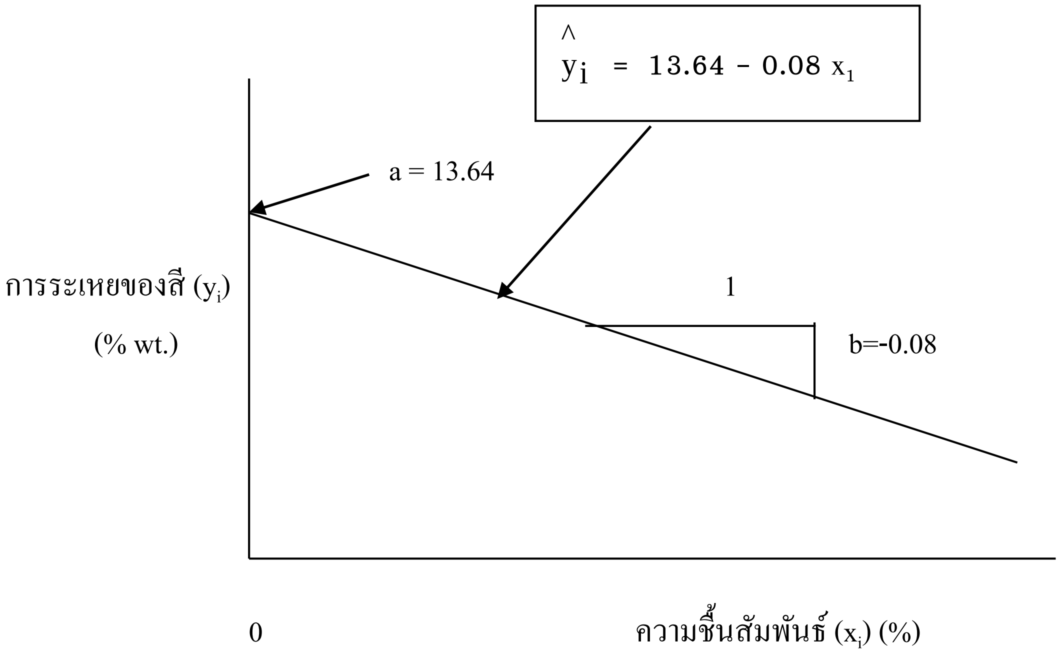
$\sum x_i^2 = 76{,}308.53 \qquad \sum y_i^2 = 2{,}286.07$

$\sum x_i\, y_i = 11{,}824.88$

$$b = \frac{25\,(11{,}824.88) - (1{,}314.9)\,(235.7)}{25\,(76{,}308.53) - (1{,}314.9)^2}$$

$$= -0.08$$

$$a = \frac{235.7}{25} - (0.08)\ \frac{(1{,}314.90)}{25} = 13.64$$



ถ้า $x_i = 50\%$ จงหา $\hat{y}_i$

$$\hat{y}_i = 13.64-0.08\ (50)$$

$$= 9.64\ \%\ \text{โดยน้ำหนัก}$$

6.3.2 คุณสมบัติของ Residual หรือ Estimation Error

$$e_i = y_i - \hat{y}_i \approx N(0,\sigma^2) \qquad \text{……………..…………….(6.5)}$$

6.3.3 การแปลงเป็นสมการเชิงเส้น (Transformation to Linearization)

(1)	y	=	ae^{bx}	Exponential
	$\ln y$	=	$\ln a + bx$	
	$\therefore\ y^*$	=	$a^* + bx$	Linear
(2)	y	=	ax^b	Power
	$\log\ y$	=	$\log a + b \log x$	
	y^*	=	$a^* + bx^*$	Linear
(3)	y	=	$a + b\ (\frac{1}{x})$	Recipocal
	y	=	$a + bx^*$	Linear

$$\text{(4)} \quad y = \frac{x}{ax+b} \qquad \text{Hyperbolic}$$

$$\frac{1}{y} = \frac{ax + b}{x}$$

$$\frac{1}{y} = a + b\left(\frac{1}{x}\right)$$

$$y^* = a + bx^* \qquad \text{Linear}$$

ตัวอย่างที่ 6.2 จากการตรวจวัดค่าความดันแก๊ส (P) ที่ปริมาตร (V) ต่าง ๆ ได้ข้อมูลดังตาราง

$V\ (cm^3)$	50	60	70	90	100
$P\ (kg/cm^2)$	64.7	51.3	40.5	25.9	7.8

ตามกฎ Ideal Gas Law ค่า P และ V มีความสัมพันธ์ดังสมการ $PV^{\lambda} = C$ จงหาค่าสัมประสิทธิ์ C และ λ

จาก $P_i V_i^{\lambda} = C$ เมื่อ $i = 1, \ldots, 5$

$\ln P_i = \ln C - \lambda \ln V_i$

ผลการแปลงค่า P_i และ V_i ด้วย ln แสดงอยู่ในตาราง

P_i	V_i	$\ln P_i$	$\ln V_i$	$\widehat{\ln P_i}$	$\hat{P}_i$	$e_i = P_i - \hat{P}_i$
64.7	50	4.16976	3.91202	4.37853	79.7	-15.0
51.3	60	3.93769	4.09434	3.89474	49.1	2.2
40.5	70	3.70130	4.24850	3.48571	32.6	7.9
25.9	90	3.25424	4.49981	2.81885	16.8	9.1
7.8	100	2.05412	4.60517	2.53928	12.7	-4.9

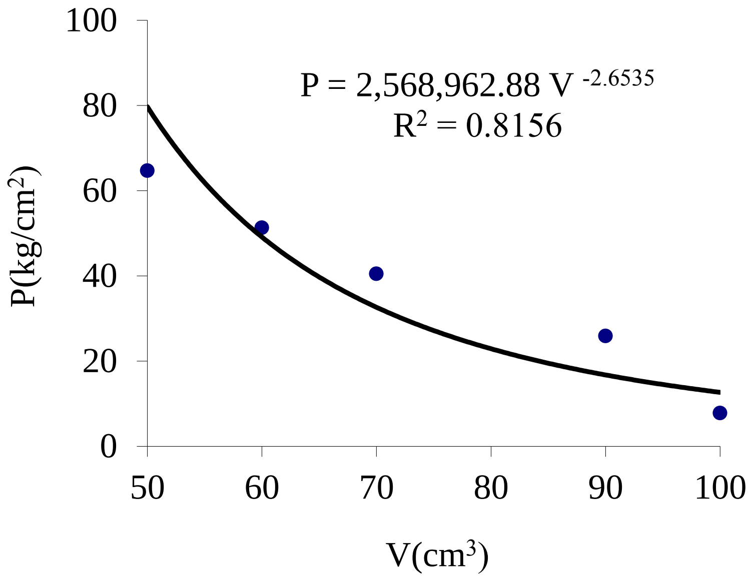
จากการวิเคราะห์ด้วยสมการถดถอยเชิงเส้นแบบง่ายพบว่า

$\ln C = 14.7589$

$C = 2{,}568{,}862.88$

$\gamma = 2.6535$

ดังนั้น $PV^{2.6535} = 2{,}568{,}862.88$



6.3.4 คุณสมบัติของ Least Square Estimates

$\mu_{y|xi} = \alpha + \beta x_i$

$Y|x_i = \alpha + \beta x_i + E_i$

$\hat{y}_i = a + bx_i$

$y_i = a + bx_i + e_i$

E_i = Independent Random Variable ($\mu_{Ei} = 0$; $\sigma^2_{Ei} = \sigma^2$)

[$Y|xi = \mu_{Y|xi} + Ei$; $E(Y|xi) = \mu_{Y|xi}$; $Var(Y|xi) = Var(Ei) = \sigma^2$]

a และ b หาจาก Sample (x_i , y_i เมื่อ i = 1, 2, ..., n)

ถ้าทำการทดลองใหม่โดยให้ x_i (i = 1, 2, ..., n) มีค่าเท่าเดิม จะได้ y_i (i = 1, 2, ..., n) ใหม่ ซึ่งต่างจาก y_i เดิม และทำให้ a, b ที่คำนวณจากแต่ละ Sample แตกต่างกันออกไป ($a \in A$; $b \in B$)

ดังนั้น A, B จึงเป็น Random Variable ซึ่งแปรผันตาม Y_i ซึ่งมี Mean $\mu_{Y|xi}$ และ $\sigma^2_{Y|xi} = \sigma^2$

$Y_i = A + Bx_i$ (Y_i, A และ B เป็น Random Variable)

$$B = \frac{n\sum x_i Y_i - (\sum x_i)(\sum Y_i)}{n\sum x_i^2 - (\sum x_i)^2}$$

$$A = \frac{\sum Y_i}{n} - B\frac{\sum x_i}{n}$$

Mean และ Variance ของตัวแปรสุ่มในสมการถดถอยแสดงอยู่ในตารางที่ 6.1

ตารางที่ 6.1 Mean และ Variance ของตัวแปรสุ่มในสมการถดถอย

Variable	Mean	Variance
x_i	$\bar{x}$	$S^2_x = \frac{Sxx}{n-1} = \frac{\sum(x_i - \bar{x})^2}{n-1}$
Y_i	$\mu_{Y\|xi}$	σ^2 ; $[S^2_y = \frac{Syy}{n-1} = \frac{\sum(y_i - \bar{y})^2}{n-1}]$
A	α	$\sigma^2_A = \frac{\sum x_i^2}{n} \cdot \frac{\sigma^2}{Sxx}$
B	β	$\sigma^2_B = \frac{\sigma^2}{Sxx}$
Ei	0	$\sigma^2_{Ei} = \sigma^2$ ให้ S^2 เป็น Estimator ของ σ^2 $S^2 = \frac{SSE}{n-2} = \frac{Syy - b\,Sxy}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2}$

6.3.5 คุณสมบัติของ Σ ที่สำคัญ

(1) $\sum (x_i - \bar{x}) = 0$ ………………..…(6.6)

(2) $\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x})\, y_i - \sum (x_i - \bar{x})\, \bar{y}$

$= \sum (x_i - \bar{x})\, y_i$

$= \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = \sum x_i y_i - n\bar{x}\bar{y}$

$= \sum x_i y_i - \sum \bar{x}\bar{y} = Sxy$ ……(6.7)

(3) $\sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x})\, x_i = \sum x^2_i - \frac{(\sum x_i)^2}{n}$

$= \sum x^2_i - n\bar{x}^2 = \sum x^2_i - \sum \bar{x}^2 = Sxx$ ..(6.8)

$$(4) \qquad \sum (y_i-\bar{y})^2 = \sum (y_i-\bar{y})y = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= \sum y_i^2 - n\bar{y}^2 = \sum y_i^2 - \sum \bar{y}^2 = Syy \quad ..(6.9)$$

6.3.6 การ Derive หา Mean และ Variance ของสัมประสิทธิ์ A และ B

$$B = \frac{\sum (x_i-\bar{x})(Y_i-\bar{Y})}{\sum (x_i-\bar{x})^2} = \frac{sxy}{sxx}$$

$$= \frac{\sum (x_i-\bar{x})\,Y_i}{\sum (x_i-\bar{x})^2} \qquad \ldots\ldots\ldots\ldots\ldots(6.10)$$

Take Expected Value

$$\mu_B = E(B) = \frac{\sum (x_i-\bar{x})\,E(Y_i)}{\sum (x_i-\bar{x})^2}$$

$$= \frac{\sum (x_i-\bar{x})}{\sum (x_i-\bar{x})^2} \cdot \mu_{Y|xi}$$

$$= \frac{\sum (x_i-\bar{x})}{\sum (x_i-\bar{x})^2} (\alpha + \beta\, xi)$$

$$= \alpha \cancelto{0}{\frac{\sum (x_i-\bar{x})}{\sum (x_i-\bar{x})^2}} + \beta \cancelto{1}{\frac{\sum (x_i-\bar{x})\,x_i}{\sum (x_i-\bar{x})^2}}$$

$$\mu_B = E(B) = \beta \qquad \ldots\ldots\ldots\ldots\ldots(6.11)$$

$$\boxed{B = \text{Unbiased Estimate of } \beta}$$

$$VAR(B) = \frac{VAR\{\sum (x_i-\bar{x})\,Y_i\}}{[\sum (x_i-\bar{x})^2]^2}$$

$$= \frac{\sum VAR(x_i-\bar{x})\,Y_i}{[\sum (x_i-\bar{x})^2]^2}$$

$$= \frac{\sum (x_i-\bar{x})^2\ VAR(Y_i)}{[\sum (x_i-\bar{x})^2]^2}$$

$$= \frac{\sigma^2 (\sum (x_i - \bar{x})^2)}{[\sum (x_i - \bar{x})^2]^2}$$

$$\sigma^2_B = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \frac{\sigma^2}{Sxx} \quad \text{………………(6.12)}$$

$$\boxed{B \sim N(\beta, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}) \sim N(\beta, \frac{\sigma^2}{Sxx})}$$

จาก $A = \bar{Y} - B\bar{x}$

$$E(A) = E(\bar{Y}) - \bar{x}E(B)$$

$$= \frac{E(Y_1) + E(Y_2) + \ldots\ldots + E(Y_n)}{n} - \bar{x}\beta \quad \ldots\ldots[E(Yi) = E(\alpha + \beta x_i + E_i)]$$

$$= \frac{(\alpha + \beta x_1) + (\alpha + \beta x_2) + \ldots\ldots + (\alpha + \beta x_n)}{n} - \bar{x}\beta$$

$$= \alpha + \beta\bar{x} - \bar{x}\beta = \alpha \quad \text{(Unbiased Estimate) …………(6.13)}$$

$$VAR(A) = VAR(\bar{Y} - B\bar{x})$$

$$= VAR\ \bar{Y} + \bar{x}^2\ VAR(B)$$

$$= \frac{\sigma^2}{n} + \bar{x}^2\ VAR(B)$$

$$= \frac{\sigma^2}{n} + \bar{x}^2 \{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}\}$$

$$= \sigma^2 \{\frac{\sum (x_i - \bar{x})^2 + n\bar{x}^2}{n \sum (x_i - \bar{x})^2}\}$$

$$= \sigma^2 \{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n} + n\bar{x}^2}{n \sum (x_i - \bar{x})^2}\}$$

$$= \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \sigma^2 \quad \text{………………(6.14)}$$

$$A \sim N\left(\alpha, \frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}\sigma^2\right) \sim N\left(\alpha, \frac{\sum x_i^2}{n} \cdot \frac{\sigma^2}{Sxx}\right)$$

$$E_i = Y_i - \mu_{Y|xi}$$

$$E(E_i) = E(Y_i) - \mu_{Y|xi} = O \qquad \text{(Unbiased Estimate)}$$

$$Var(E_i) = Var(Y_i) + O = \sigma^2$$

ให้ S^2 เป็นตัวประมาณค่าของ σ^2

$$S^2 = \frac{SSE}{n-2} = \frac{\sum(y_i - y_i)^2}{n-2} \qquad \text{(Model Error Variance)}$$

$$\begin{aligned} SSE &= \sum (y_i - a - bx_i)^2 \\ &= \sum (y_i - (\bar{y} - b\bar{x}) - bx_i)^2 \\ &= \sum [\, (y_i - \bar{y}) - b(x_i - \bar{x})\,]^2 \\ &= \sum [\, (y_i - \bar{y})^2 - 2b\,(x_i - \bar{x})\,(y_i - \bar{y}) + b^2\,(x_i - \bar{x})^2] \\ &= Syy - 2b\,Sxy + b^2\,Sxx \end{aligned}$$

$$\boxed{\text{จาก } b = \frac{Sxy}{Sxx}}$$

$$\begin{aligned} SSE &= Syy - 2b\,Sxy + \frac{Sxy}{Sxx}\,.\,b.\,Sxx \\ &= Syy - b\,Sxy \qquad \ldots\ldots\ldots\ldots\ldots(6.15) \end{aligned}$$

$$S^2 = \frac{Syy - b\,Sxy}{n-2} \approx \sigma^2_{Ei} \text{ (Degree of Freedom = n-2)} \ldots\ldots\ldots(6.16)$$

6.3.7 Inferences Concerning Regression Coefficient

การอนุมานเกี่ยวกับสัมประสิทธิ์สมการถดถอยทำได้ 2 แบบคือ

- การทดสอบนัยสำคัญ (Significance)
- การหาช่วงความเชื่อมั่น (Confidence Interval)

$$\left.\begin{aligned} &N\left(A\,;\,\alpha, \sigma^2_A = \frac{\sum x_i^2}{n} \cdot \frac{\sigma^2}{Sxx}\right) \\ &N\left(B\,;\,\beta, \sigma^2_B = \frac{\sigma^2}{Sxx}\right) \end{aligned}\right\} \leftarrow N(Y_i : \mu_{Y|xi}, \sigma^2)$$

(1) การทดสอบนัยสำคัญของ β

Standard Normal $Z = \dfrac{B-\beta}{\sigma_B}$

$= \dfrac{B-\beta}{\sigma/\sqrt{Sxx}}$

Student's t $T = \dfrac{Z}{\sqrt{\dfrac{\chi^2}{\upsilon}}} = \dfrac{Z}{\sqrt{\dfrac{\chi^2}{n-2}}}$

Chi – Square $\chi^2 = \dfrac{(n-2)\,S^2}{\sigma^2}$

D.F. = n-2 = n-m-1 = n-1-1 (m = 1 เนื่องจากใช้ b เป็นตัวประมาณค่าของ B)

$$T = \frac{B-\beta}{\dfrac{\sigma}{\sqrt{Sxx}}\sqrt{\dfrac{(n-2)\,S^2}{\sigma^2\,(n-2)}}} \quad \text{……..(6.17)}$$

$$= \frac{B-\beta}{\dfrac{S}{\sqrt{Sxx}}}$$

(2) T มีการแจกแจงแบบ Students't ที่ (n-2) degree of freedom

$\mu_{Y|x} = \alpha+\beta x$

ทดสอบสมมติฐานว่า $\mu_{Y|x}$ เปลี่ยนตามค่า x หรือไม่ หรือ $\beta=0$ หรือ $\beta = \beta o$

Ho : $\beta = \beta o$

H_i : $\beta > \beta o$ (ทดสอบทางเดียวด้านบน)

หรือ H_1 : $\beta < \beta o$ (ทดสอบทางเดียวด้านล่าง)

หรือ H_1 : $\beta \neq \beta o$ (ทดสอบสองทาง)

$t = \dfrac{b-\beta o}{\dfrac{S}{\sqrt{Sxx}}}$ (Statistic)

ตัวอย่างที่ 6.3 จากตัวอย่างที่ 6.1

Y = อัตราการระเหยของสี

x = ค่าความชื้นสัมพันธ์

$\hat{y}$ = 13.64-0.08x

จงทดสอบนัยสำคัญของ β (b ~ β ~-0.08) ที่ระดับนัยสำคัญ (α) = 1% และ 5%

H_o = $\beta = \beta o = 0$

H_1 = $\beta \neq \beta o \neq 0$ (Two tails)

$$t = \frac{b-\beta}{\frac{S}{\sqrt{Sxx}}} = \frac{b-0}{\frac{S}{\sqrt{Sxx}}}$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 76{,}308.83 - \frac{1{,}314.9^2}{25}$$

= 7,150.05

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 2{,}286.07 - \frac{235.7^2}{25}$$

= 63.89

$$S_{xy} = \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= 11{,}824.44 - \frac{1{,}314.9 \text{ x } 235.7}{25}$$

= -572.44

SSE = $S_{yy} - b\, S_{xy}$

= 63.89 - (-0.08) (-572.44)

= 18.09

$$S^2 = \frac{SSE}{n-2} = \frac{18.09}{25-2} = 0.79$$

$$t = \frac{-0.08-0}{\frac{\sqrt{0.79}}{\sqrt{7{,}150.05}}} = -7.61$$

จากตาราง Students't (n-2) = 2.3

$t_{\alpha/2,\, n-2}$ = 2.069 (α = 5%)

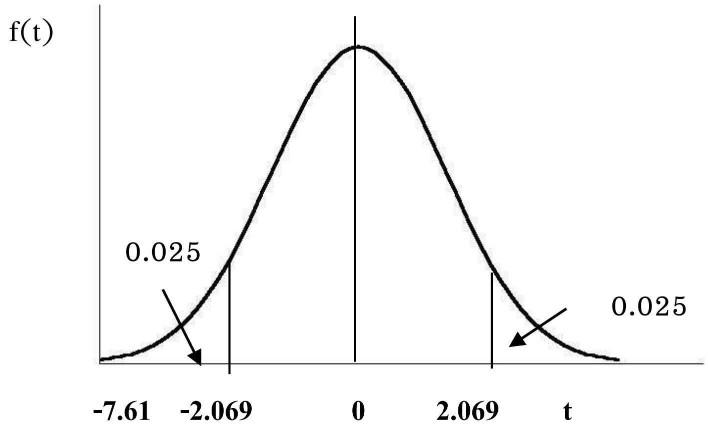
$t_{\alpha/2,\, n-2}$ = 2.807 (α = 1%)

$|t = -7.61| > t_{\frac{\alpha}{2},\, n-2}$

Reject H_o : $\beta = 0$

Accept H_1 : $\beta \neq 0$

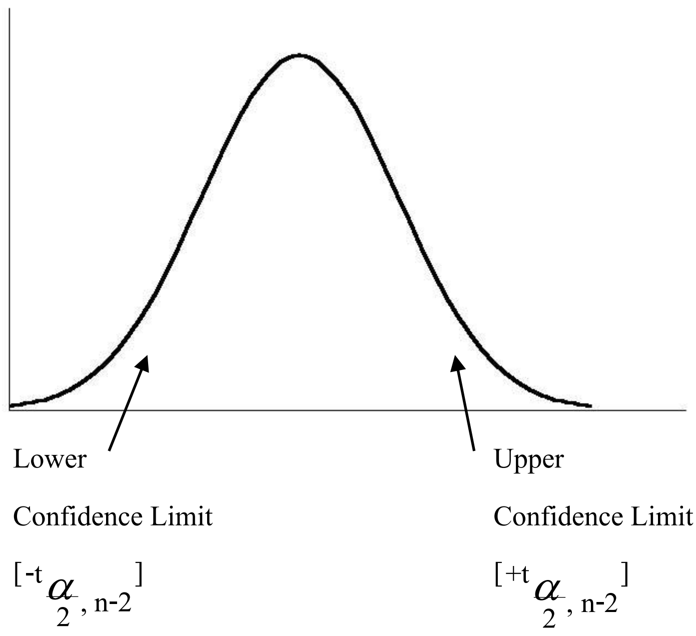
$(\beta < 0)$



(3) การหาช่วงความเชื่อมั่นของ β

$$P\left[-t_{\frac{\alpha}{2},\upsilon} \leq t = \frac{b-\beta}{\frac{S}{\sqrt{Sxx}}} < t_{\frac{\alpha}{2},\upsilon}\right] = 1-\alpha$$

$$b-t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}} \leq \beta < b+t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}} \qquad \text{..............................(6.18)}$$



ตัวอย่างที่ 6.4 จากตัวอย่างที่ 6.1 และ 6.3 จงหา Confidence Interval ของ β ที่ Confidence Level 99% หรือ α = 1%

$$b = -0.08 \ ; \ t_{0.005,23} = 2.807$$

$$-0.08 - \frac{2.807\sqrt{0.79}}{\sqrt{7{,}150.05}} < \beta < -0.08 + \frac{2.807\sqrt{0.79}}{\sqrt{7{,}150.05}} \quad -0.109 < \beta < -0.051$$

แสดงว่า $\beta \neq 0 \rightarrow \beta < 0$

(4) การทดสอบนัยสำคัญของ α

$$N\left(A : \alpha, \sigma^2_A = \frac{\sum x_i^2}{n} \cdot \frac{\sigma^2}{Sxx}\right)$$

Standard Normal Z

$$Z = \frac{A-\alpha}{\sqrt{\frac{\sum x_i^2}{n} \cdot \frac{\sigma^2}{Sxx}}} = \frac{A-\alpha}{\sigma\sqrt{\frac{\sum x_i^2}{n.\ Sxx}}} \quad \text{.....(6.19)}$$

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{\upsilon}}} = \frac{Z}{\sqrt{\frac{\chi^2}{n-2}}}$$

$$= \frac{A-\alpha}{\sigma\sqrt{\frac{\sum x_i^2}{n.\ Sxx}}\sqrt{\frac{(n-2)\,S^2}{\sigma^2\,(n-2)}}}$$

$$= \frac{A-\alpha}{\frac{S}{\sqrt{Sxx}}\sqrt{\frac{\sum x_i^2}{n}}} \quad \text{.....(6.20)}$$

= Students't ที่ (n-2) degree of freedom

การทดสอบสมมติฐาน

Null Hypothesis [Ho : $\alpha = \alpha o$]

Alternative Hypothesis [H_1 : $\alpha \neq \alpha o$]

$$t = \frac{a-\alpha}{\frac{S}{\sqrt{Sxx}}\sqrt{\frac{\sum x_i^2}{n}}}$$

ตัวอย่างที่ 6.5 จากตัวอย่างที่ 6.1 $\hat{y}$ = 13.64-0.08x จงทดสอบว่า $\alpha = \alpha_0 = 13$ ที่ α = 10% หรือ $\alpha = 0.1$

$$H_o \quad : \quad \alpha = 13 \qquad\qquad (a = 13.64)$$

$$H_1 \quad : \quad \alpha \neq 13$$

$$t = \frac{13.64 - 13}{\dfrac{\sqrt{0.79}}{\sqrt{7{,}150.05}}\sqrt{\dfrac{76{,}308.53}{25}}}$$

$$= 1.10 < t_{\frac{\alpha}{2}, n-2} = t_{0.05, 23} = 1.714$$

Accept H_o : $\alpha = 13$ ที่ Confidence Level 90 %

(5) การหา Confidence Interval ของ α

$$P\left[-t_{\frac{\alpha}{2},\upsilon} \leq t = \frac{a - \alpha}{\dfrac{S}{\sqrt{Sxx}}\sqrt{\dfrac{\sum x_i^2}{n}}} < t_{\frac{\alpha}{2},\upsilon}\right] = 1-\alpha$$

$$a - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{Sxx}}\sqrt{\frac{\sum x_i^2}{n}} \leq \alpha < a + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}}\sqrt{\frac{\sum x_i^2}{n}} \qquad \text{.........(6.21)}$$

ตัวอย่างที่ 6.6 จากตัวอย่างที่ 6.1 จงหา Confidence Interval ของ α ที่ Confidence Level 90% (α = 10%)

$$t_{\frac{\alpha}{2}, n-2} = t_{0.05, 23} = 1.714$$

$$13.64 - 1.714\frac{\sqrt{0.79}}{\sqrt{7{,}150.5}}\sqrt{\frac{76{,}308.53}{25}} < \alpha < 13.64 + 1.714\frac{\sqrt{0.79}}{\sqrt{7{,}150.5}}\sqrt{\frac{76{,}308.53}{25}}$$

$$12.61 < \alpha < 14.64$$

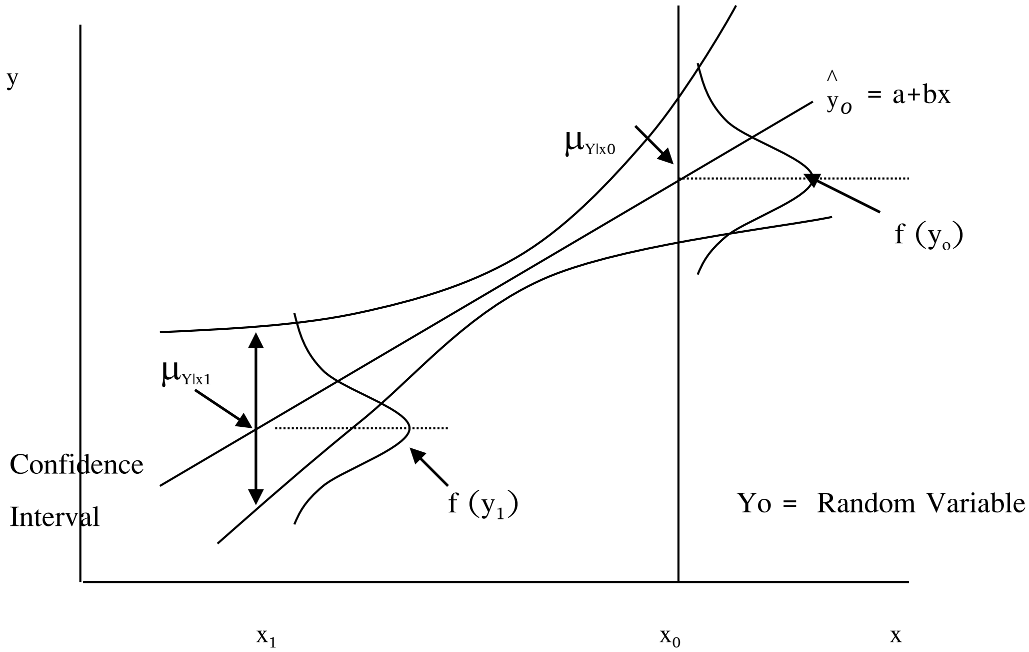
Accept H_o : $\alpha = 13$

6.3.8 การคาดคะเน (Prediction)

ถ้า $x = x_o$ ใด ๆ

$$\hat{y}_o = a + bx_o$$

y_o คือค่าประเมิน (Estimator) ของ $\mu_{Y|xo}$ หรือ μ_{Yo}



ที่ $x = x_o$

$$\hat{Y}_o = A + Bx_o$$

$$E(\hat{Y}_o) = E(A) + x_o E(B)$$

$$\mu_{\hat{Y}_o} = \alpha + x_o \cdot \beta$$

$$= \mu_{Y|xo}$$

$$\hat{Y}_o = A + Bx_o$$

$$= \bar{Y} - B\bar{x} + Bx_o$$

$$VAR(\hat{Y}_o) = VAR[\bar{Y} + (x_o - \bar{x})B]]$$

$$= VAR(\bar{Y}) + (x_o - \bar{x})^2 \; VAR(B)$$

$$= \frac{\sigma^2}{n} + (x_o - \bar{x})^2 \; \frac{\sigma^2}{Sxx}$$

$$= \left(\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}\right)\sigma^2$$

$$N\left(\hat{Y}o : \mu_{Y|xo}, \left(\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}\right)/\sigma^2\right.$$

Standard Normal

$$Z = \frac{\hat{Y}o - \mu_{Y|xo}}{\sigma\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.22)$$

$$T = \frac{\hat{Y}o - \mu_{Y|xo}}{\sigma\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}} \cdot \frac{S}{\sigma}}$$

$$= \frac{\hat{Y}o - \mu_{Y|xo}}{S\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.23)$$

Confidence Interval ของ $\mu_{Y|xo}$

$$P\left[-t_{\frac{\alpha}{2},\upsilon} \le t = \frac{\hat{Y}o - \mu_{Y|xo}}{S\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}} < t_{\frac{\alpha}{2},\upsilon}\right] = 1-\alpha$$

$$\hat{Y}o - t_{\frac{\alpha}{2}} . S\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}} \le \mu_{Y|xo} < \hat{Y}o + t_{\frac{\alpha}{2}} . S\sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}} \quad \ldots\ldots\ldots(6.24)$$

ตัวอย่างที่ 6.7 จากตัวอย่างที่ 6.1

$\hat{y}_i$ = 13.64 - 0.08 x_I

ถ้า x_I = x_o = 50%

$\hat{y}_o$ = 13.64 - 0.08 x 50

= 9.64 % โดยน้ำหนัก

จงหา Confidence Interval ของ $\mu_{Y/xo}$ ที่ α = 10 %

$$t_{\frac{\alpha}{2}, n-2} = t_{0.05, 23} = 1.714$$

$$S^2 = 0.79, \quad Sxx = 7,150.05, \quad n = 2.5$$

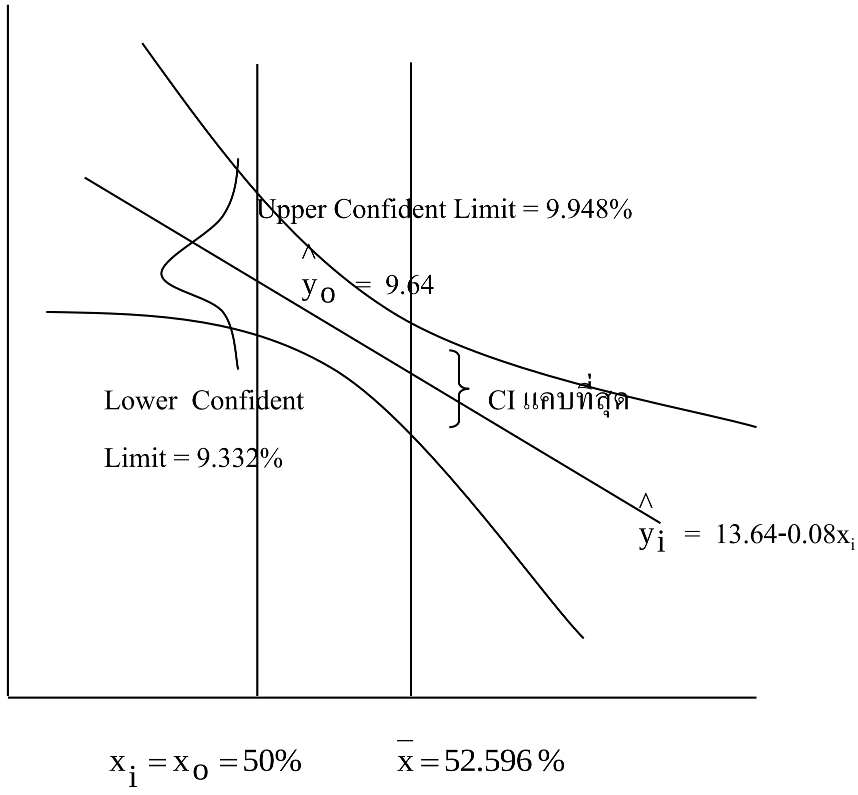
$$\bar{x} = \frac{\sum x_i}{n} = \frac{1,314.9}{2.5} = 52.596$$

$$9.64 - 1.714\sqrt{0.79}\sqrt{\frac{1}{25} + \frac{(50 - 52.586)}{7,150.05}}$$

$$< \mu_{Y/xo} < 9.64 + 1.714\sqrt{0.79}\sqrt{\frac{1}{25} + \frac{(50 - 52.596)^2}{7,150.05}}$$

$$9.64 - 0.3083 < \mu_{Y|xo} < 9.64 + 0.3083$$

$$9.332 < \mu_{Y|xo} < 9.948$$



ตัวอย่างที่ 6.8 ในการศึกษาอัตราการใช้น้ำมันของรถยนต์ที่วิ่งนอกเมือง รถยนต์จากโรงงานที่ผลิตในปีเดียวกันแต่ต่างรุ่น จำนวน 10 คัน มาทดสอบวิ่งในระยะทาง 1,000 กม. บันทึกอัตราการใช้น้ำมัน เป็น กม./ลิตร (y) และน้ำหนักรถยนต์เป็นตัน (x) ได้ผลดังตาราง

i	1	2	3	4	5	6	7	8	9	10
y	17.9	16.5	16.4	16.8	18.8	15.5	17.5	16.4	15.9	18.3
x	1.35	1.9	1.7	1.8	1.3	2.05	1.6	1.8	1.85	1.40

จงหาช่วงความเชื่อมั่นของค่าพยากรณ์เฉลี่ย $\mu_{Y/xo}$

$n = 10, \quad \sum x_i = 16.75, \quad \sum y_i = 170$

$\sum x_i^2 = 28.6375, \quad \sum y_i^2 = 2900.46$

$\sum x_i y_i = 282.408$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= 28.6375 - \frac{(16.75)^2}{10}$$

$$= 0.581$$

$$S_{xy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 2900.46 - \frac{(170)^2}{10}$$

$$= 10.46$$

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= 282.405 - \frac{(16.75)(170)}{10}$$

$$= -2.345$$

$$\hat{y}_i = a + bx_i$$

$$b = \frac{Sxy}{Sxx} = \frac{-2.345}{0.581} = -4.03$$

$$a = \bar{y} - b\bar{x}$$

$$= \frac{170}{10} - (-4.03)\left(\frac{16.75}{10}\right)$$

$$= 23.75$$

$$\hat{y}_i = 23.75 - 4.03x_i$$

Confidence Interval ของ β และ α ที่ $(1-\alpha) = 99\ \%$

$$t_{\frac{\alpha}{2}, n-2} = t_{0.005, 8} = 3.355$$

$$S^2 = \frac{SSE}{n-2}$$

$$= \frac{Syy - b.Sxy}{n - 2}$$

$$= \frac{10.46 - (-4.03)(-2.345)}{8}$$

$$= 0.126$$

$$b - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}} < \beta < b + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}}$$

$$-4.03 - 3.355 \frac{\sqrt{0.126}}{\sqrt{0.581}} < \beta < -4.03 + 3.355 \frac{\sqrt{0.126}}{\sqrt{0.581}}$$

$$-4.03 - 1.56 < \beta < -4.03 + 1.56$$

$$-5.59 < \beta < -2.47$$

ที่ Confidence Level 99%

สรุป $\beta \neq 0$ และ $\beta < 0$ ที่ $\alpha = 1\%$

$$a - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}} \sqrt{\frac{\sum x_i^2}{n}} < \alpha < a + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{Sxx}} \sqrt{\frac{\sum x_i^2}{n}}$$

$$23.75 - 3.355 \frac{\sqrt{0.126}}{\sqrt{0.581}} \sqrt{\frac{28.6375}{10}} < \alpha < 23.75 + 3.355 \frac{\sqrt{0.126}}{\sqrt{0.581}}$$

$$23.75 - 2.63 < \alpha < 23.75 + 2.63$$

$$21.12 < \alpha < 26.38$$

สรุป $\alpha \neq 0$ และ $\alpha > 0$ ที่ $\alpha = 1\%$

ถ้า $x_i = x_0 = 1.7$ ตัน

$$\hat{y}_0 = 23.75 - 4.03x_o$$

$$= 23.75 - 4.03\,(1.7)$$

$$= 16.899 \text{ กม./ลิตร}$$

หา Confidence Limit ของ $\mu_{Y/xo}$ ที่ $(1-\alpha) = 90\%$

$$t_{\frac{\alpha}{2}, n-2} = t_{0.05, 8} = 1.86$$

$$t_{\frac{\alpha}{2}} . S \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}$$

$$= 1.86\sqrt{.126} \sqrt{\frac{1}{10} + \frac{(1.7 - 16.75/10)^2}{0.581}}$$

$= \quad 0.2099$

$$16.899 - 0.2099 < \mu_{Y/xo} < 16.899 + 0.2099$$

$$16.689 < \mu_{Y/xo} < 17.109$$

6.3.9 Prediction Interval (PI)

PI = Single Predicted Value $\pm\, t_{\frac{\alpha}{2}} \sqrt{\text{variance of } Y_o \text{ from the true } Y_o}$ ซึ่งต่างจาก

CI = Single Predicted Value $\pm\, t_{\frac{\alpha}{2}} \sqrt{\text{variance of the estimate } Y_o \text{ from } \mu_{Y|xo}}$

$\hat{Yo} = a + bx_o$ (Estimate)

$\mu_{Y|xo} = \alpha + \beta x_o$ (Mean)

$Y_o = \alpha + \beta x_o + E_o$ (True Yo)

Random Error

$\hat{Yo} - y_o$ = Error (Residual)

$\hat{Yo} - Y_o$ = Random Variable

$$\mu_{\hat{Yo} - Y_o} = E(\hat{Yo} - Y_o)$$

$$= E(\hat{Yo}) - E(Y_o)$$

$$= \mu_{Y|xo} - \mu_{Y|xo} = 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.25)$$

$$VAR(\hat{Yo} - Y_o) = VAR(\hat{Yo}) + VAR(Y_o)$$

$$= \left(\frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}\right)\sigma^2 + \sigma^2$$

$$= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}\right] \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.26)$$

$$N\left(\hat{Yo} - Y_o : 0, \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}\right]\right)$$

Standard Normal

$$Z = \frac{(\hat{Yo} - Y_o) - \mu_{\hat{Y_o} - Y_o}}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}}$$

$$T = \frac{(\hat{Y}o - Y_o) - \mu_{\hat{Y}_o - Y_o}}{S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}}$$

$$T = \frac{\hat{Y}_o - Y_o}{S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.27)$$

Prediction Interval of Y_o at α%

$$P[-t_{\frac{\alpha}{2},\upsilon} \leq t = \frac{\hat{Y}o - y_o}{S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}}} < t_{\frac{\alpha}{2},\upsilon}] = 1 - \alpha$$

$$\hat{Y}o - t_{\frac{\alpha}{2}} . S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}} < y_o < \hat{Y}o + t_{\frac{\alpha}{2}} . S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{Sxx}} \quad \ldots\ldots\ldots\ldots(6.28)$$

ตัวอย่างที่ 6.9 จากตัวอย่างที่ 6.8 จงหา Prediction Interval ของการพยากรณ์อัตราการใช้น้ำของรถยนต์ ซึ่งมีน้ำหนัก 1.7 ตัน ถ้า α = 10%

$$\hat{y}_o = 23.75 - 4.03\, x_o$$

$$\hat{y}_o = 23.75 - 4.03\,(1.7) = 16.899 \text{ กม./ลิตร}$$

$$t_{\frac{\alpha}{2}, n-2} = t_{0.05, 8} = 1.86$$

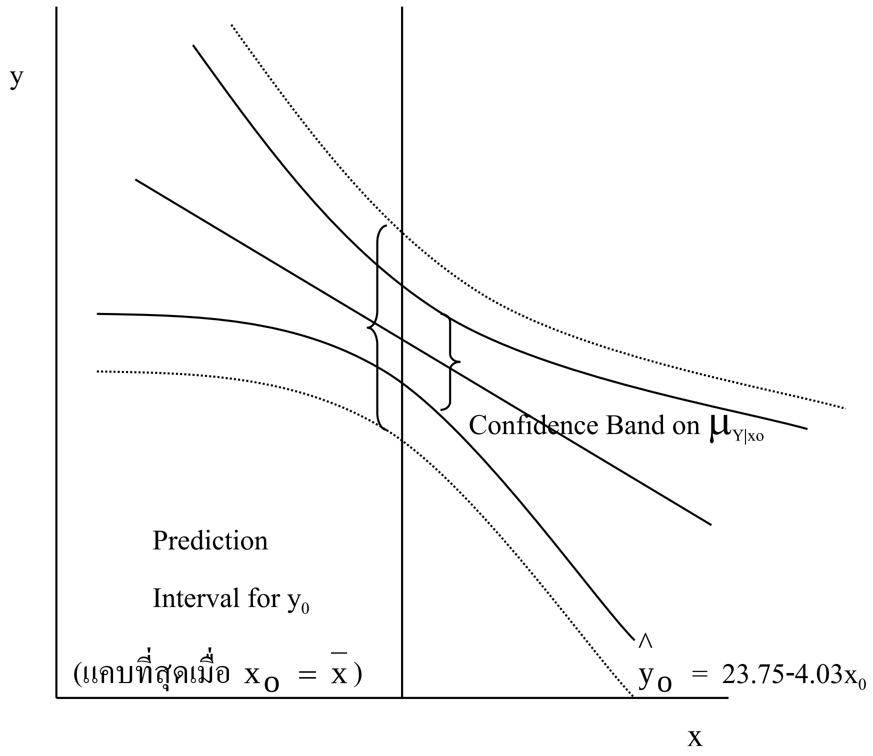
$$t_{0.05, 8}\, S\sqrt{1 + \frac{1}{n} + \frac{x_o - \bar{x})^2}{Sxx}}$$

$$= 1.86\sqrt{0.126}\ \sqrt{1 + \frac{1}{10} + \frac{(1.7 - 1.675)^2}{0.581}}$$

$$= 0.6927$$

$$16.899 - 0.6927 < y_o < 16.899 + 0.6927$$

$$16.206 < y_o < 17.592$$



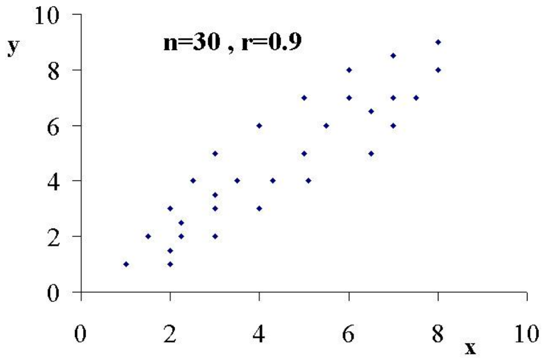
6.4 สหสัมพันธ์เชิงเส้น (Linear Correlation)

6.4.1 Coefficient of Correlation

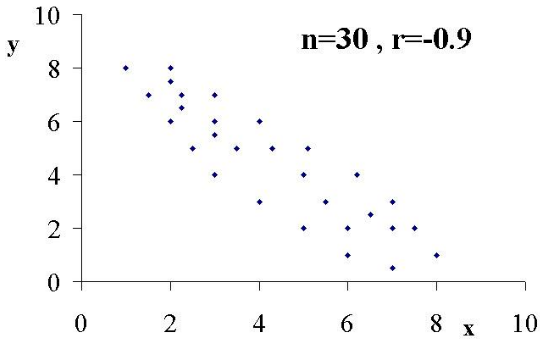
ค่าสัมประสิทธิ์ที่ใช้วัดความสัมพันธ์ระหว่าง Y และ X เรียกว่าสัมประสิทธิ์สหสัมพันธ์ Coefficient of Correlation (ρ) ดังรูปที่ 10.3

$$\rho = \frac{COV\,(x, y)}{\sqrt{VAR\,(x)\ VAR\,(y)}} = \frac{Sxy}{\sqrt{Sxx.\ Syy}} \quad \ldots\ldots\ldots\ldots(6.29)$$

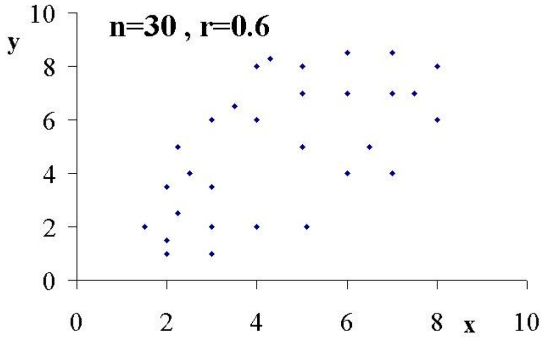
r คือตัวประมาณค่าของ ρ



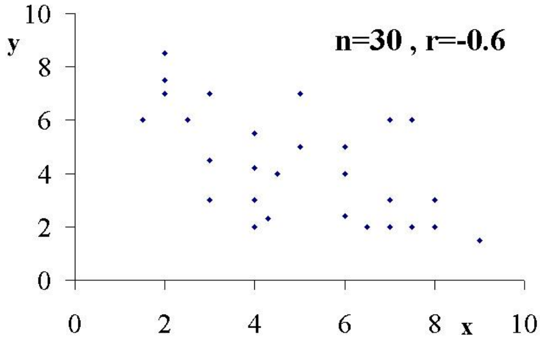
(1) Positive Correlation



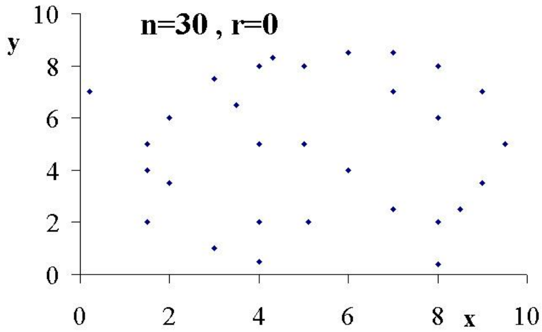
(4) Negative Correlation



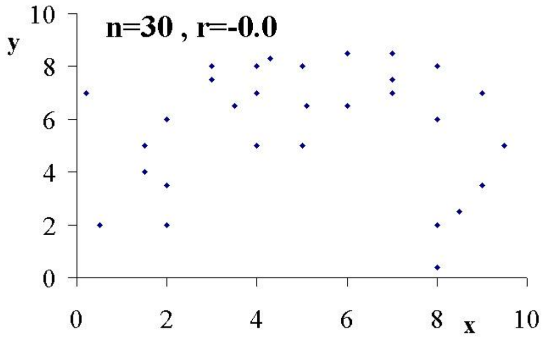
(2) Positive Correlation May Be Present



(5) Negative Correlation May Be Present



(3) No Correlation



(6) No Correlation

รูปที่ 6.3 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร x และ y กรณีต่าง ๆ

$$\mathrm{COV}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{Sxy}{n-1} \qquad \text{……..…………………..(6.30)}$$

$$\mathrm{VAR}(X) = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{Sxx}{n-1}$$

$$VAR(Y) = \frac{\sum(y_i - \bar{y})^2}{n-1} = \frac{Syy}{n-1}$$

$$\rho = r = \frac{Sxy}{\sqrt{Sxx.\ Syy}} \quad (-1 < r < 1)$$

$$r^2 = \frac{S_{xy}^2}{Sxx.\ Syy} \quad \text{(Coefficient of Determination)} \quad \ldots\ldots\ldots\ldots\ldots\ldots(6.31)$$

จาก $b = \frac{Sxy}{Sxx}$

$$SSE = Syy\text{-}b\ Sxy$$

$$\frac{SSE}{Syy} = 1\text{-}b\ \frac{Sxy}{Sxx}$$

$$= 1\text{-}\ \frac{Sxy}{Sxx} \cdot \frac{Sxy}{Syy}$$

$$= 1\text{-}r^2$$

$$r^2 = 1 - \frac{SSE}{Syy} \approx 0\text{-}1 \quad \ldots\ldots\ldots\ldots\ldots\ldots(6.32)$$

ถ้า $SSE = 0$

$r^2 = 1$ แสดงว่า X และ Y มีความสัมพันธ์เชิงเส้นอย่างสมบูรณ์

ถ้า $SSE = Syy$

$r^2 = 0$ No Correlation

ตัวอย่างที่ 6.10 ในการศึกษาผลกระทบของน้ำเสียในทะเลสาบแห่งหนึ่ง ได้มีการวัดปริมาณความเข้มข้นของสารประกอบไนเตรท โดยใช้วิธีเดิมและวิธีใหม่ เปรียบเทียบกัน ถ้าพบว่าความเข้มข้นของสารประกอบไนเตรทจากวิธีเดิม และวิธีใหม่มีความสัมพันธ์เชิงเส้นสูง จะใช้วิธีใหม่วัดค่า

I	1	2	3	4	5	6	7	8	9	10
x วิธีเดิม	25	40	120	75	150	300	270	400	450	575
y วิธีใหม่	30	80	150	80	200	350	240	320	470	583

$n = 10$ $\quad \sum x_i = 2,405 \quad \sum x_i^2 = 900,775$

$\sum y_i = 2,503 \quad \sum y_i^2 = 919,489$

$\sum x_i y_i = 902,475$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= 900{,}775 - \frac{(2405)^2}{10} = 322{,}372.5$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= 919{,}489 - \frac{(2503)^2}{2} = 292{,}988.7$$

$$S_{xy} = \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= 902{,}475 - \frac{(2{,}405)(2{,}503)}{10}$$

$$= 300{,}503.5$$

$$r = \frac{Sxy}{\sqrt{Sxx.\ Syy}}$$

$$= \frac{300{,}503.5}{\sqrt{322{,}372.5 \text{ x } 292{,}988.7}}$$

$$= 0.978$$

แสดงว่า 97.8% ของความแปรปรวนของ Y เป็นผลมาจากสมการถดถอย ถ้าสามารถหาสมการถดถอย y และ x จะสามารถประมาณ y จาก x ได้

6.4.2 Bivariate Normal Distribution

X, Y = Normal Random Variable

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-\frac{1}{2}\left(\frac{x - \mu x}{\sigma x}\right)^2} \quad -\infty < x < \infty$$

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma y} e^{-\frac{1}{2}\left(\frac{y - \mu y}{\sigma y}\right)^2} \quad -\infty < y < \infty$$

ถ้า X และ Y Independent

$$f(x, y) = f(x)\, f(y)$$

$$= \frac{1}{2\pi\ \sigma x\ \sigma y} e^{-\frac{1}{2}\left[\left(\frac{x - \mu x}{\sigma x}\right)^2 + \left(\frac{y - \mu y}{\sigma y}\right)^2\right]} \qquad \text{...........(6.33)}$$

ถ้า X และ Y ไม่ Independent

$$f(x, y) = f(x)\, f(y/x)$$

$$= \frac{1}{2\pi\,\sigma x\,\sigma y\sqrt{1-\rho^2}} \text{Exp}\ [-\frac{1}{2(1-\rho^2)}\{(\frac{y-\mu y}{\sigma y})^2$$

$$-2\rho(\frac{x-\mu x}{\sigma x})(\frac{y-\mu y}{\sigma y}) + (\frac{x-\mu x}{\sigma x})^2\}] \qquad \text{.....................(6.34)}$$

เมื่อ $-1<\rho<1$; $-\infty<x, y, <\infty$; $\sigma x > 0$; $\sigma y > 0$

ถ้า $\rho = 0$ (X, Y Independent)

สมการ (10.33) = สมการ (10.34)

เนื่องจาก $$\rho^2 = 1-\frac{\sigma^2}{\sigma y^2} = \beta^2\frac{\sigma x^2}{\sigma y^2} \qquad \text{............(6.35)}$$

σy และ σx มีค่าเป็น +

ρ และ β จะมีเครื่องหมายเหมือนกัน

จึงสามารถทดสอบสมมติฐานได้ว่า

Ho : $\rho = 0$

หรือ Ho : $\beta = 0$

โดยใช้ค่าสถิติ

$$t(n-2) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

เมื่อ $$r = \frac{Sxy}{\sqrt{Sxx.\ Syy}} = b\sqrt{\frac{Sxx}{Syy}}$$

$$b = \frac{Sxy}{Sxx}$$

ตัวอย่างที่ 6.11 จากตัวอย่างที่ 10.10 จงหาความสัมพันธ์ระหว่างวิธีใหม่ และวิธีเดิม

พบว่า $r = 0.978 \rightarrow \rho \neq 0$

เนื่องจาก $n = 10$: มีค่าน้อยมากจึงควรทดสอบ r

H_0 : $\rho = 0$

H_1 : $\rho \neq 0$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$= \frac{0.978\sqrt{8}}{\sqrt{1-0.978^2}}$$

$$= 13.26$$

จากตาราง $t_{0.0005,\,8} = 5.041$

$t > t_{0.0005,\,8}$

Reject H_o : สรุป $\rho \neq 0$ ที่ระดับความเชื่อมั่น 99.9 %

6.4.3 หา Confidence Limits ของ ρ (Pearson Correlation Coefficient)

$$\frac{1}{2}\ \ln\left(\frac{1+r}{1-r}\right) = \text{Normal Random Variable}$$

$$\mu = \frac{1}{2}\ \ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$\sigma^2 = \frac{1}{n-3}$$

$$Z = \frac{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}}$$

$$P\left[-Z_{\frac{\alpha}{z}} \le \frac{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}} < Z_{\frac{\alpha}{z}}\right] = 1-\alpha$$

$$-Z_{\frac{\alpha}{z}} \le \frac{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}} < Z_{\frac{\alpha}{z}}$$

Lower Confidence Limit ของ ρ

$$= \frac{(1+r)-(1-r)\exp\left(2Z_{\frac{\alpha}{z}} / \sqrt{N-3}\right)}{(1+r)+(1-r)\exp\left(2Z_{\frac{\alpha}{z}} / \sqrt{N-3}\right)} \qquad \text{………………(6.36)}$$

Upper Confidence Limit ของ ρ

$$= \frac{(1+r)-(1-r)\exp\left(-2Z_{\frac{\alpha}{z}} / \sqrt{N-3}\right)}{(1+r)+(1-r)\exp\left(-2Z_{\frac{\alpha}{z}} / \sqrt{N-3}\right)} \qquad \text{………………(6.37)}$$

ตัวอย่างที่ 6.12 จงหา Confidence Limit ของ ρ จากตัวอย่างที่ 6.10 เมื่อ α = 5%

$$z_{0.025} = 1.96$$

Lower Limit

$$= \frac{(1+0.978)-(1-0.978)\,\text{Exp}\,(2*1.96/\sqrt{7})}{(1+0.978)-(1-0.978)\,\text{Exp}\,(2*1.96/\sqrt{7})}$$

$$= \frac{1.978 - 0.022 \text{ x } 4.4}{1.978+0.022 \text{ x } 4.4} = \frac{1.881}{2.075} = 0.907$$

Upper Limit

$$= \frac{(1+0.978)-(1-0.978)\,\text{Exp}\,(-2*1.96/\sqrt{7})}{(1+0.978)+(1-0.978)\,\text{Exp}\,(-2*1.96/\sqrt{7})}$$

$$= 0.995$$

$0.907 < \rho < 0.995$ ที่ Confidence Level 95 %

6.5 Multiple Linear Regression

(การถดถอยเชิงเส้นพหุคูณ)

$$Y = f(x1, x2, \ldots\ldots\ldots, xn)$$

K Independent Variable ($x_1, x_2, \ldots\ldots\ldots, x_n$)

$$\mu_{Y|x1, x2, \ldots\ldots, xk} = \beta_0+\beta_1 x_1+\beta_2 x_2+\ldots\ldots\ldots+\beta_k x_k$$

$$y = b_0+b_1x_1+b_2x_2+\ldots\ldots\ldots\ldots+b_kx_k$$

<u>Sample</u> $\{ (x_{1i}, x_{2i},\ldots\ldots,x_{ki}, y_i) ; i=1,\ldots\ldots\ldots n\}$ เมื่อ $n > k$

ที่ค่า $x_{1i}, x_{2i},\ldots\ldots\ldots\ldots x_{ki}$ ใด ๆ

$$y_i = b_0+b_1x_{1i}+b_2x_{2i}+\ldots\ldots+b_kx_{ki}+e_I \quad \ldots\ldots\ldots\ldots\ldots\ldots(3.38)$$

$$\hat{y}_i = b_0+b_1x_{1i}+b_2x_{2i}+\ldots\ldots\ldots b_kx_{ki} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6.39)$$

$$\boxed{y_i - \hat{y}_i = e_i}$$

6.5.1 Least Square Method

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)$$

$$= \sum (y_i-b_0-b_1x_{1i}-b_2x_{2i}-b_kx_{ki})^2$$

$$\frac{\partial SSE}{\partial b_0} = 0$$

$$\frac{\partial SSE}{\partial b_1} = 0$$

.

.

$$\frac{\partial SSE}{\partial b_k} = 0 \qquad ....(k+1)$$

$$\sum (y_i-b_0-b_1x_{1i}-b_2x_{2i}..................-b_kx_{ki}) = 0$$

$$\sum x_{1i} (y_i-b_0-b_1x_{1i}-b_2x_{2i}....................-b_kx_{ki}) = 0$$

$$\sum x_{2i} (y_i-b_0-b_1x_{1i}-b_2x_{2i}....................-b_kx_{ki}) = 0$$

$$\sum x_{ki} (y_i-b_0-b_1x_{1i}-b_2x_{2i}.....................-b_kx_{ki}) = 0 \qquad \ldots\ldots\ldots\ldots(6.40)$$

$$\sum y_i = bo\sum 1_0+b_1\sum x_{1i}+b_2\sum x_{2i}+..........................+b_k\sum x_{ki})$$

$$\sum x_{1i}\, y_i = b_0\sum x_{1i}+b_1\sum x^2_{1i}+b_2\sum x_{1i}x_{2i}+....................+b_k\sum x_{1i}x_{ki})$$

$$\sum x_{2i}\, y_i = b_0\sum x_{2i}+b_1\sum x_{2i}x_{1i}+b_2\sum x^2_{2i}+....................+b_k\sum x_{2i}x_{ki})$$

$$\sum x_{ki}\, y_i = b_0\sum x_{ki}+b_1\sum x_{ki}x_{1i}+b_2\sum x_{ki}x_{2i}+.................+b_k\sum x^2_{ki}) \ldots\ldots\ldots(6.41)$$

แก้ k+1 สมการ หาพารามิเตอร์ b_0..............b_k = (k+1) พารามิเตอร์

กรณีที่มีตัวแปรอิสระ (X_i) มาก การแก้สมการโดยวิธีธรรมดา (Gaussian Elimination) จะทำได้ลำบาก

6.5.2 วิธี Matrix

สมการ Multiple Linear Regression in Matrix Form

$$Y = X\beta + \in \qquad \ldots\ldots\ldots\ldots\ldots\ldots(6.42)$$

$$\underset{[nx1]}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix}$$

$$\underset{[n\ x\ (k+1)]}{X} = \begin{array}{cccc} \mathbf{1} & \mathbf{X_1.} & \mathbf{X_2.} & \mathbf{X_k.} \\ \end{array} \begin{bmatrix} 1 & X_{11} & X_{21} \ldots\ldots\ldots\ldots & X_{k1} \\ 1 & X_{12} & X_{22} \ldots\ldots\ldots\ldots & X_{k2} \\ 1 & X_{13} & X_{23} \ldots\ldots\ldots\ldots & X_{k3} \\ & \text{-------------------------} & & \\ 1 & X_{1n} & X_{2n} \ldots\ldots\ldots\ldots & X_{kn} \end{bmatrix}$$

$$\underset{[(k+1)*1]}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

$$\underset{(nx1)}{\in} = \begin{bmatrix} \in_1 \\ \in_2 \\ . \\ . \\ . \\ \in_n \end{bmatrix}$$

$\in$ = Error in Estimation

= Independent Normally Distributed

$\sim \; N(\in ; 0, \; \sigma^2 \, In)$

In = Identity Matrix

ให้ Matrix $\hat{\beta} \sim \beta$

$$\beta = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_k \end{bmatrix}$$

ซึ่งจะหาได้ดังนี้

$$Y = X\hat{\beta} + \in$$

$$X^T Y = X^T X \hat{\beta} + X^T \in \qquad (X^T = \text{X-transpose})$$

ให้ $A = X^T X$ ……………………….(6.43)

$g = X^T Y$ ……………………….(6.44)

$A\hat{\beta} = g$ ……………………….(6.45)

$\hat{\beta} = A^{-1} g = (X^T X)^{-1} X^T Y$ ……………………….(6.46)

$$A = X^T X$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_{11} & x_{21} & x_{13} \ldots\ldots\ldots\ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} \ldots\ldots\ldots\ldots & x_{2n} \\ & ------------ & & \\ x_{k1} & x_{k2} & x_{k3} \ldots\ldots\ldots\ldots & x_{kn} \end{bmatrix}_{(k+1) \times n} \begin{bmatrix} 1 & x_{11} & x_{21} \ldots\ldots\ldots x_{k1} \\ 1 & x_{12} & x_{22} \ldots\ldots\ldots x_{k2} \\ 1 & x_{13} & x_{23} \ldots\ldots\ldots x_{k3} \\ & ------------ & \\ 1 & x_{1n} & x_{2n} \quad x_{kn} \end{bmatrix}_{n \times (k+1)}$$

6.5.3 พิสูจน์ว่า $\hat{\beta}$ เป็น Unbiased Estimator ของ β

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= (X^T X)^{-1} X^T (X\beta + \in)$$

$$= (X^T X)^{-1} (X^T X)\beta + (X^T X)^{-1} X^T \in$$

$$= \beta + (X^T X)^{-1} X^T \in \qquad \text{………………….……..(6.47)}$$

$$E(\hat{\beta}) = E(\beta) + (X^T X)^{-1} X^T E(\in) \qquad \text{………………….……..(6.48)}$$

$$= \beta \leftarrow \text{(Unbiased)}$$

VARIANCE ของ β

$$VAR(\hat{\beta}) = E(\hat{\beta} - \beta)^2$$

จาก $\hat{\beta} - \beta = (X^TX)^{-1}X^T \in$

$$VAR(\hat{\beta}) = E[\{(X^TX)^{-1}X^T \in\} \{(X^TX)^{-1}X^T \in\}]$$

จาก $(X^TX)^{-1}X^T \in \;=\; \in^T X(X^TX)^{-1}$

$$VAR(\hat{\beta}) = E\{(X^TX)^{-1}X^T \in . \in^T X(X^TX)^{-1}\}$$

$$= (X^TX)^{-1}X^T E(\overset{\sigma^2}{\cancel{\in \in^T}}) X (X^TX)^{-1}$$

$$= \sigma^2 \{(X^TX)^{-1} \underset{In}{\cancel{X^TX}}\} (X^TX)^{-1}$$

$$= \sigma^2 \, In \, (X^TX)^{-1} \qquad \text{………………….……..(6.49)}$$

ค่าประมาณของ σ^2 หาได้จากตาราง ANOVA

Sources	D.F.	SS	MS	F
Regression (R)	k	$\beta^T X^T Y - n\bar{y}^2$	$MSR = \frac{SSR}{k}$	$\frac{MSR}{MSE} = f$
Error (E)	n-k-1	$Y^TY - \beta^T X^T Y$ (Syy-b Sxy)	$MSE = \frac{SSE}{n-k-1}$	
Total (T)	n-1	$Y^TY - n\bar{y}^2$		

$$S^2 = MSE \approx \sigma^2$$

$$= \frac{Y^T Y - \hat{\beta}^T x^T y}{n-k-1}$$

ค่า f ใช้ทดสอบนัยสำคัญของ Multiple Linear Regression

ถ้า $f > f_{\alpha,(k,n-k-1)}$ ให้ Reject Ho

หรือสรุปว่าสมการ Multiple Linear Regression สามารถใช้อธิบายค่าจริงของ y (true situation) ได้อย่างเพียงพอ อย่างมีนัยสำคัญทางสถิติที่ α %

ตัวอย่างที่ 6.13 ให้ y = ยอดขายสินค้าชนิดหนึ่ง (แสนบาท)

x_1 = ค่าใช้จ่ายในการโฆษณาทางวิทยุ

x_2 = ค่าใช้จ่ายในการโฆษณาทางหนังสือพิมพ์

ผลการสุ่มตัวอย่างได้ข้อมูลดังตาราง จงวิเคราะห์หาความสัมพันธ์ระหว่าง y และ x_1, x_2

i	y	x_1	x_2	yx_1	yx_2	$x_1 x_2$
1	7	4	1	28	7	4
2	12	7	2	84	24	14
3	17	9	5	153	85	45
n = 4	20	12	8	240	160	96
$\sum$	56	32	16	505	276	159
SS	882	290	94			

$$A = X^T X = \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} \sum x_{2i} \\ \sum x_{2i} & \sum x_{1i} \sum x_{2i} & \sum x_{2i}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 32 & 16 \\ 32 & 290 & 159 \\ 16 & 159 & 94 \end{bmatrix}$$

$$g = X^T Y = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix} = \begin{bmatrix} 56 \\ 505 \\ 276 \end{bmatrix}$$

$4b_0 + 32b_1 + 16b_2 = 56$ ……………………(1)

$32b_0 + 290b_1 + 159b_2 = 505$ ……………………(2)

$16b_0 + 159b_1 + 94b_2 = 276$ ……………………(3)

แก้สมการโดยวิธี Matrix

$$\hat{\beta} = A^{-1} g$$

$$= (X^T X)^{-1} X^T Y$$

$$a_{ij}^{-1} = \frac{(-1)^{i+j}\ dij}{|A|}$$

$|A|$ = Determinant of A

d_{ij} = Determinant of Sub-matrix formed by Eliminating the i-row and j-column of A

$$|A| = 4\,(290 \times 94 - 159 \times 159) - 32(32 \times 94 - 16 \times 159) + 16(32 \times 159 - 16 \times 290)$$

$$= 4 \times 1979 - 32 \times 464 + 16 \times 448$$

$$= 236$$

$$A^{-1} = \frac{1}{236}\begin{bmatrix} 1979 & -464 & 448 \\ -464 & 120 & -124 \\ 448 & -124 & 136 \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{236}\begin{bmatrix} 56 \\ 505 \\ 276 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6441 \\ 1.6610 \\ 0.0169 \end{bmatrix}$$

$\hat{y} = 0.6441 + 1.6610\, x_1 + 0.0169\, x_2$

Analysis of Variance (ANOVA)

$$\hat{\beta}^T X^T Y = [0.6441 \quad 1.6610 \quad 0.0169] \begin{bmatrix} 56 \\ 505 \\ 276 \end{bmatrix}$$

$$= 879.55$$

$$n\bar{y}^2 = 4\left(\frac{56}{4}\right)^2 = 784$$

$$Y^T Y = [7 \quad 12 \quad 17 \quad 20] \begin{bmatrix} 7 \\ 12 \\ 17 \\ 20 \end{bmatrix}$$

$$= \sum y_i^2$$

$$= 882$$

SSR = Sum of Square of Regression

$$= \hat{\beta}^T X^T Y - n\bar{y}^2$$

$$= 879.55 - 784 = 95.55$$

$$SSE = Y^T Y - \hat{\beta}^T X^T Y$$

$$= 882 - 872.55 = 2.45$$

$$SST = Y^T Y - n\bar{y}^2 = 882 - 784 = 98$$

SOURCE	D.F.	SS	M.S	F
Regression (R)	2	95.55	47.75	$\frac{47.75}{2.45}$
Error (E)	1	2.45	2.45	= 19.5
Total	3	98		

r^2 = SSR / SST = 95.55 / 98 = 0.975

r = 0.987

6.6 แบบฝึกหัด

1. ในการศึกษาเกี่ยวกับความชื้นของส่วนผสมของสินค้าชนิดหนึ่ง, X และความหนาแน่นของสินค้าชนิดนี้ที่ผลิตได้ (Y) ได้ข้อมูลดังต่อตาราง

X	Y
5	7.4
6	9.3
7	10.6
10	15.4
12	18.1
15	22.2
18	24.1
20	24.8

ก. จงเขียนแผนภาพกระจาย

ข. จากกราฟในข้อ (ก) X และ Y มีแนวโน้มที่จะมีความสัมพันธ์เชิงเส้นตรงหรือไม่

ค. จงประมาณสมการถดถอย $\mu_{Y/x} = \alpha + \beta x$

ง. จงเขียนสมการปกติจากข้อมูลชุดนี้

2. ในการศึกษาความสัมพันธ์ระหว่างความต้านทานแรงดึงของผลิตภัณฑ์กระดาษชนิดหนึ่ง กับปริมาณไม้เนื้อแข็งที่ผสมในเยื่อกระดาษ ได้ทำการทดลองการเก็บตัวอย่าง 10 ตัวอย่างจากโรงงานที่ผลิตได้ข้อมูลดังต่อไปนี้

แรงดึง (tensile strength)	ปริมาณไม้เนื้อแข็ง (%)
160	10
171	13
175	15
182	18
184	18.5
181	20
188	22
193	25
195	28
200	30

ก. จงประมาณสมการเส้นถดถอย $\mu_{Y/x} = \alpha + \beta x$

ข. จงประมาณความต้านทานต่อแรงดึง เมื่อปริมาณไม้เนื้อแข็งเป็น 20% และ 25%

ค. จงประมาณค่าความผิดพลาดในข้อ (ข)

ง. จงทดสอบสมมติฐาน $\beta = 0$ โดยใช้ระดับนัยสำคัญ 0.05 แย้งกับสมมติฐาน $\beta \neq 0$

จ. จงหาช่วงความเชื่อมั่น 95% ของความต้านทานต่อแรงดึงเฉลี่ย เมื่อปริมาณไม้เนื้อแข็งที่ผสมในเยื่อกระดาษเป็น 30%

3. ในการศึกษาผลกระทบของอัตราเร็วในการคนสี ในกระบวนการเคมีชนิดหนึ่งกับปริมาณสิ่งเจือปนที่ตกค้างในสีนั้น ได้ข้อมูลดังนี้

อัตราเร็วในการคน (rpm) (X)	ปริมาณสิ่งเจือปนในสี (%) (Y)
20	8.4
22	9.5
24	11.8
26	10.4
28	13.3
30	14.8
32	13.2
34	14.7
36	16.4
38	16.5
40	18.9
42	18.5

$n = 12, \quad \sum_i x_i = 372, \quad \sum_i y_i = 166.4, \quad \bar{x} = 31$

$\bar{y} = 13.87, \quad \sum_i x_i^2 = 12{,}104, \quad \sum_i y_i^2 = 2{,}435.14, \quad \sum_i x_i y_i = 5{,}419.60$

ก. จงประมาณสมการถดถอย $\mu_{Y/x} = \alpha + \beta x$

ข. จงทดสอบสมมติฐาน $\beta = 1.5$ แย้งกับสมมติฐาน $\beta < 1.5$ โดยใช้ระดับนัยสำคัญ 0.05

ค. จงหาค่าประมาณแบบจุดของความแปรปรวนของปริมาณสิ่งเจือปนในสี

ง. ด้วยความเชื่อมั่น 95% จงหาช่วง $\mu_{Y/xo}$ เมื่อ $x_o = 26$ rpm

จ. จงสร้าง 95% ช่วงความเชื่อมั่นของค่าพยากรณ์ Y เมื่อ $x_o = 26$ rpm

4. ในการพิจารณาความสัมพันธ์ระหว่างแรงเค้น (S) กับจำนวนรอบการใช้งานก่อนที่จะใช้งานไม่ได้ (N) ของโลหะผสมชนิดหนึ่งว่าเป็นไปตามสมการ

$$S = A * N^{-m}$$

โดยที่ A และ m เป็นค่าคงที่ที่ไม่ทราบค่า ได้เก็บรวบรวมข้อมูลไว้ดังต่อไปนี้

แรงดัน (psi x 10^3)	**N** (จำนวนรอบ x 10^6 รอบ)
55.0	0.223
50.5	0.925
43.5	6.750
42.5	18.100
42.0	29.100
41.0	50.500
35.7	126.000
34.5	215.000
33.0	445.000
32.0	420.000

จงหาค่าประมาณของ A และ m

ข้อแนะนำ : สมการ $S = A * N^{-m}$ เขียนใหม่ในรูปของ

$\log S = \log A - m \log N$ กำหนดให้ $S = Y$, $\log A = \alpha$, $-m = \beta$, $\log N = X$

5. กำหนดให้ Y คือ ระดับคะแนนเฉลี่ย (GPA) ของนิสิตชั้นปีที่ 1 หลังจากสิ้นปีการศึกษาแรกและ X คือ คะแนนสอบเข้ามหาวิทยาลัยของนิสิตคนนั้น สุ่มตัวอย่างนิสิตชั้นปีที่ 1 มา 20 คน พร้อมรวบรวมคะแนนสอบเข้ามหาวิทยาลัยได้ข้อมูลนี้

$$\sum_i x_i = 100, \quad \sum_i y_i = 50, \quad \sum_i x_i^2 = 509.12$$

$$\sum_i y_i^2 = 134.84, \quad \sum_i x_i y_i = 257.66$$

ก. เส้นถดถอย $\mu_{Y/x} = \alpha + \beta x$ เมื่อ α และ β คือพารามิเตอร์ประมาณด้วยเส้นตรง $\hat{y} = a + bx$ เมื่อ a และ b เป็นค่าประมาณของ α และ β ตามลำดับ จงเขียนสมการปกติ (โดยไม่ต้องแสดงวิธีทำ)

ข. จงประมาณเส้นถดถอยดังกล่าว และหาค่า S^2

ค. จงสร้างช่วงความเชื่อมั่น 95 % สำหรับ β

ง. จงหาค่าประมาณของการเปลี่ยนแปลงของระดับคะแนนเฉลี่ย เมื่อคะแนนสอบเข้ามหาวิทยาลัยของนิสิตเพิ่มขึ้น 1 หน่วย

จ. จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ และสร้างช่วงความเชื่อมั่น 95% ของ ρ

6. ในการศึกษาความสัมพันธ์ระหว่างความหนาแน่นของประชากรผู้อยู่อาศัย และอัตราการเกิดการโจรกรรมในเมือง ๆ หนึ่ง นักวิจัยได้เก็บรวบรวมข้อมูลในปีที่ผ่านมาได้ข้อมูลดังนี้

จำนวนผู้อยู่อาศัย, X_i (คน : หน่วยพื้นที่)	อัตราการเกิดโจรกรรม, Y_i (จำนวนครั้ง : 100,000 คน)
59	209
49	180
75	195
54	192
78	215
56	197
60	208
82	189
69	123
83	201
88	214
94	212
47	205
65	186

ก. จงหาค่าประมาณของอัตราการเกิดโจรกรรมในเมืองนี้โดยเฉลี่ยในปีที่ผ่านมาขณะที่ X_o = 60 และหาค่าความผิดพลาด (e)

ข. จงหาค่าประมาณแบบจุดของ σ^2

ค. จงทดสอบสมมติฐาน $\rho = 0$ แย้งกับสมมติฐาน $\rho > 0$ โดยใช้ระดับนัยสำคัญ 0.05

ง. จงเขียนสมการปกติ (ไม่ต้องแสดงวิธีทำ)

7. ในการศึกษาความสัมพันธ์ระหว่างความดันโลหิต (Y) กับอายุ (X) ของเด็กชายอายุระหว่าง 5 ถึง 13 ปี ได้ข้อมูลดังต่อไปนี้

X	5	8	11	7	13	12	12	6
Y	63	67	74	64	75	69	90	60

ก. จงหาค่าสัมประสิทธิ์สหสัมพันธ์ (r) และทดสอบสมมติฐาน $\rho = 0.8$ แย้งกับสมมติฐาน $\rho < 0.8$

ข. จงสร้างช่วงความเชื่อมั่น 95% ของค่าพยากรณ์ y_o เมื่อ $x_o = 10$

8. ในการศึกษาความสัมพันธ์ระหว่างความต้านทานต่อแรงดึงของเส้นใยชนิดหนึ่ง (Y) กับเปอร์เซ็นต์ของฝ้ายในเส้นใยนี้ (X_1) และเวลาที่ใช้ในการอบแห้งเส้นใยนี้ (X_2) จึงได้ทำการเก็บตัวอย่างเส้นใยที่ผลิตได้ภายในสภาวะต่าง ๆ กัน 10 แบบ ได้ข้อมูลดังนี้

Y = แรงดึง	X_1 = เปอร์เซ็นต์ของฝ้าย	X_2 = เวลาที่ใช้ในการอบแห้ง
213	13	2.1
220	15	2.3
216	14	2.2
225	18	2.5
235	19	3.2
218	20	2.4
239	22	3.4
243	17	4.1
233	16	4.0
240	18	4.3

ก. จงเขียนสมการปกติ (ไม่ต้องแสดงวิธีทำ)

ข. จงประมาณสมการถดถอยพหุคูณ $\mu_{Y/x_1, x_2} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

9. ในการศึกษาความสัมพันธ์ระหว่างผลที่ได้จากกระบวนการเคมีหนึ่ง (Y) กับความเข้มข้นของตัวเร่งปฏิกิริยาทางเคมี (X_1) และอุณหภูมิขณะทำปฏิกิริยา (X_2) ได้ข้อมูลดังนี้

ผลที่ได้ (Y)	ความเข้มข้น (X_1)	อุณหภูมิ $^\circ$C (X_2)

81	1.0	150
89	1.0	180
83	2.0	150
91	2.0	180
97	1.0	150
87	1.0	180
84	2.0	150
90	2.0	180

ก. จงเขียนสมการปกติ

ข. จงประมาณสมการถดถอยพหุคูณ

ค. จงทดสอบว่า ความเข้มข้นของตัวเร่งปฏิกิริยา มีผลต่อการเปลี่ยนแปลงของผลที่ได้จากกระบวนการเคมีนี้หรือไม่ ใช้ระดับนัยสำคัญ 0.05

ง. จงประมาณค่าเฉลี่ยของผลที่ได้จากกระบวนการเคมี เมื่อทำการทดลองที่อุณหภูมิ 175°C และความเข้มข้นของตัวเร่งปฏิกิริยาเป็น 2.0

10. ในการความสัมพันธ์ระหว่างอัตราการฆ่าตัวตายกับขนาดของประชากร และอัตราการหย่าร้าง นักวิจัยได้ทำการเก็บรวบรวมข้อมูลจากเมืองต่าง ๆ 8 เมืองได้ข้อมูลดังตาราง

เมือง	จำนวนประชากร, X_1 (x 1,000 คน)	อัตราการหย่าร้าง, X_2 (ต่อ 100,000 คน)	อัตราการฆ่าตัวตาย, Y (ต่อ 100,000 คน)
Akron, Chio	679	30.4	11.6
Anaheim, CA	1,420	34.1	16.1
Buffalo, NY	1,349	17.2	9.3
Austin, Texas	296	26.8	9.1
Chicago Ill.	6,975	29.1	8.4
Columbia, SC	323	18.7	7.7
Dctroit, MI	4,200	32.6	11.3
Gary, Indiana	633	32.5	8.4

จากข้อมูลดังกล่าว นักวิจัยผู้นี้ต้องการสร้างสมการถดถอยพหุคูณเชิงเส้นจากต้นแบบ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

และได้ทำการคำนวณค่าต่าง ๆ ในรูปของเมตริกไว้ดังนี้

$$X'Y = \begin{bmatrix} 81.9 \\ 159,832 \\ 2,335.3 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 2.7831 & 1.707\text{x}10^{-5} & -0.09727 \\ 1.707\text{x}10 & 2.6961\text{x}10^{-8} & -2.55\text{x}10^{-6} \\ -0.09727 & -2.55\text{x}10^{-6} & 0.003698 \end{bmatrix}$$

ก. จงเขียนสมการปกติ

ข. จงประมาณสมการถดถอยพหุคูณดังกล่าว

ค. จงหาค่าประมาณแบบจุดของ σ^2

ง. จงประมาณอัตราการฆ่าตัวตายโดยเฉลี่ย ถ้าเมือง ๆ หนึ่งมีประชากรอาศัยอยู่ 2,500,000 คน และอัตราการหย่าร้างในเมืองนี้เป็น 27.8 ต่อแสนคน

จ. ถ้าเมือง 2 เมืองมีประชากรอาศัยอยู่เท่ากัน ถ้าอัตราการหย่าร้าง (ต่อแสนคน) ของประชากรใน 2 เมืองนี้ต่างกัน 1 หน่วย จงประมาณความแตกต่างของอัตราการฆ่าตัวตายของประชากรในเมือง 2 เมืองนี้

บทที่ 7
การวางแผนการทดลองและการวิเคราะห์ความแปรปรวน
(Experimental Design and Analysis of Variance)

7.1 คำนำ

การวิเคราะห์ความแปรปรวน (ANOVA) คือวิธีการวิเคราะห์ผลการทดลอง เพื่อการทดสอบสมมติฐานของ Mean และ Variance ของประชากรหลาย ๆ ชุด (k ชุด) พร้อม ๆ กัน เช่น

	k Random Variables หรือ k Populations					
	X_1	X_2	....	X_i	....	X_k
คุณสมบัติของประชากร	μ_1, σ_1	μ_2, σ_2		μ_i, σ_i		μ_k, σ_k
คุณสมบัติของตัวอย่าง	$\bar{x}_1, S_1, n_1$	$\bar{x}_2, S_2, n_2$		$\bar{x}_i, S_i, n_i$		$\bar{x}_k, S_k, n_k$

สมมติฐานในการทดสอบค่าเฉลี่ย

$H_o : \mu_1 = \mu_2 = ... = \mu_i = ... = \mu_k$

$H_1 : \mu_i$ ไม่เท่ากัน หรืออย่างน้อย 1 คู่ไม่เท่ากัน (At least two of the means are not equal)

- ถ้าใช้วิธีทดสอบสมมติฐานในบทที่ 8 ต้องทดสอบถึง $^{K}C_2 = \dfrac{k\,!}{2\,!\,(k-2)\,!}$
- ข้อดีของ ANOVA คือ สามารถแบ่ง Total Variation หรือ Variance ออกเป็นส่วน ๆ เช่น
    - Between Treatment Variation (Systematic/Aggregate)
    - Within Treatment Variation (Random)

Total Variation (SST) = Between Treatment Variation (SSA)+Within Treatment Variation (SSE)

หรือ = Systematic Variation + Random Variation

เทอม SS_ คือ Sum of Square ซึ่งมีค่าเท่ากับ $\sum_{i=1}^{n}(x_i - \bar{x})^2$ เมื่อหารด้วย n-1 จะได้ Variance

7.2 การวางแผนการทดลองทางสถิติ

ถ้าแบ่งตามวิธีการสุ่มจะแบ่งออกได้เป็น 3 วิธี ดังนี้

CRD = Completely Randomized Design (แผนการทดลองแบบสุ่มอย่างสมบูรณ์)

RCB = Randomized Complete Block Design (แผนการทดลองแบบสุ่มตัวอย่างสมบูรณ์ในบล๊อค)

LS = Latin Square Design (แผนการทดลองแบบละตินสแควร์)

ถ้าแบ่งตามจำนวนปัจจัยจะแบ่งได้เป็น 1 ปัจจัย (One Factor) และหลายปัจจัย (More Than One Factors) กรณีหลายปัจจัยจะแบ่งออกได้เป็น Factorial, Split Plot และ Strip Plot

7.2.1 การวางแผนการทดลองกรณี 1 ปัจจัย

กรณีมี 4 สิ่งทดลอง ๆ ละ 4 ซ้ำ (4 treatments @ 4 replications) สามารถวางแผนการทดลองแบบ CRD, RCB และ LS ได้ดังนี้

A	D	B	C	A
B	C	A	D	B
B	A	C		
D	C	D		

Experimental Unit (หน่วยทดลอง)

Treatment (สิ่งทดลอง)

CRD (ไม่มีข้อจำกัดในการสุ่มจัดสิ่งทดลองลงในหน่วยทดลอง เนื่องจากแต่ละหน่วยทดลองไม่แตกต่างกัน)

BLOCK				
1	A	D	C	B
2	C	A	B	D
3	C	D	B	A
4	B	C	A	D

RCB (มีข้อจำกัด 1 ด้าน คือแต่ละ Block จะต้องสุ่มจัดสิ่งทดลอง 1 ชุด(A, B, C, D) เนื่องจากแต่ละหน่วยทดลองในแต่ละบล๊อคไม่แตกต่างกัน)

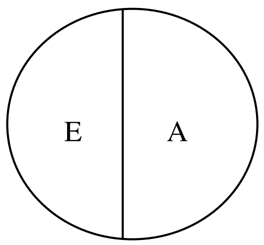
COLUMN

ROW	1	2	3	4
1	A	C	B	D
2	B	D	A	C
3	C	A	D	B
4	D	B	C	A

LS (มีข้อจำกัด 2 ด้าน คือแต่ละ Column และแต่ละ Row จะสุ่มจัดสิ่งทดลอง 1 ชุด (A, B, C, D) เนื่องจากหน่วยทดลองในแนว Column/Row อาจไม่เหมือนกัน)

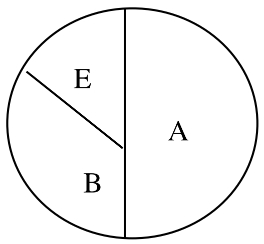
แหล่งความแปรปรวนสำหรับกรณีต่าง ๆ แสดงดังรูปที่ 7.1 ซึ่งจะเห็นได้ว่า

$$SSE\ (CRD) > SSE\ (RCB) > SSE\ (LS) \qquad \text{……………………………(7.1)}$$



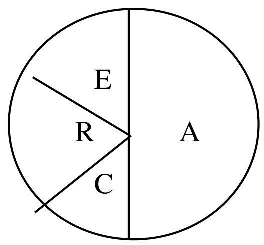
กรณี CRD

SST = SSA+SSE



กรณี RCB

SST = SSA+SSB+SSE



กรณี LS

SST = SSA+SSC+SSR+SSE

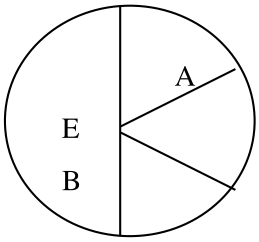
(A = treatment ; E = Error ; B = Block ; R = Row ; C = Column)

รูปที่ 7.1 แหล่งความแปรปรวนกรณีมี 1 ปัจจัย ซึ่งมีการวางแผนการทดลองแบบ CRD, RCB และ LS

7.2.2 การวางแผนการทดลองกรณีหลายปัจจัย

(1) การวางแผนการทดลองแบบ Factorial

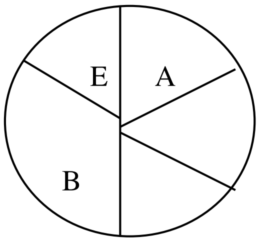
ใช้ศึกษาปัจจัยตั้งแต่ 2 ปัจจัยขึ้นไป โดยให้ความสำคัญทุกปัจจัยเท่าๆ กัน กรณีมี 2 ปัจจัย (A และ B) แหล่งความแปรปรวนจะสามารถแยกออกได้ดังรูปที่ 7.2



กรณี CRD

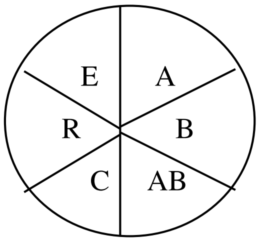
SST = SSA+SSB+SS(AB)+SSE

SS(AB) คือ Sum of Square ของปฏิกิริยาสัมพันธ์ระหว่างปัจจัย A และ B



กรณี RCB

SST = SSA+SSB+SS(AB)+SS(Block)+SSE



กรณี LS

SST = SSA+SSB+SS(AB)+SSR+SSC+SSE

(A, B = ปัจจัย ; AB = ปฏิกริยาสัมพันธ์ระหว่างปัจจัย A และ B
R = Row ; C = Column)

รูปที่ 7.2 แหล่งความแปรปรวนของการวางแผนการทดลองแบบ Factorial กรณี 2 ปัจจัย

(2) การวางแผนการทดลองแบบ Split Plot

ใช้ศึกษาปัจจัยตั้งแต่ 2 ปัจจัยขึ้นไป คล้ายกับการทดลองแบบ Factorial แต่การทดลองแบบ Factorial ให้ความสำคัญต่อปัจจัยที่ทำการศึกษาเท่า ๆ กัน ซึ่งผิดกับแผนการทดลองแบบ Split-Plot ที่ให้ความสำคัญของบางปัจจัยมากกว่าปัจจัยอื่น ๆ โดยการจัดปัจจัยที่ให้ความสำคัญมากเป็นแปลงย่อย (Sub Plot) และปัจจัยที่ให้ความสำคัญน้อยเป็นแปลงใหญ่ (Main Plot) โดยอาจจะใช้แผนการทดลองแบบ CRD, RCB และ LS ก็ได้ ข้อเสียที่สำคัญของ Split Plot คือสามารถศึกษาอิทธิพลของ Main Plot

ได้อย่างหยาบ ๆ เท่านั้น ความแปรปรวนของการทดลองแบบ Split Plot สามารถแบ่งเป็นส่วนต่าง ๆ ได้ดังสมการ

กรณี CRD

SST = SSA+SSE(A)+SSB+SSE(B)+SS(AB) ….………..(7.2)

กรณี RCB

SST = SSA+SSE(A)+SSB+SSE(B)+SS(AB)+SS(Block) ……………………..(7.3)

กรณี LS

SST = SSA+SSE(A)+SSB+SSE(B)+SS(AB)+SSR+SSC ……………………..(7.4)

เมื่อ

SSE(A), SSE(B) = Sum of Square of Error ของปัจจัย A และ B ตามลำดับ

(3) การวางแผนการทดลองแบบ Strip Plot

ใช้ทดสอบปัจจัย 2 ปัจจัย สามารถวัดปฏิกิริยาสัมพันธ์ (Interaction) ได้แน่นอนและแม่นยำมากกว่าอิทธิพลของปัจจัยหลัก โดยกำหนดให้ปัจจัยแรกเป็นปัจจัยในแนวนอน (Horizontal Factor) ปัจจัยที่ 2 เป็นปัจจัยในแนวดิ่ง (Vertical Factor) ส่วน Intersection Plot เป็นแปลงที่ใช้วัดปฏิกิริยาสัมพันธ์ระหว่างระดับต่าง ๆ ของปัจจัยทั้ง 2 ดังแสดงในรูปที่ 7.3

(Vertical Factor)

B3	B1	B4	B5	B2	
					A2
					A4
					A3
					A1

(Horizontal Factor)

รูปที่ 7.3 ผังการทดลองแบบ Strip-Plot

ข้อดีของ Strip-Plot เปรียบกับ Split-Plot คือ Strip-Plot ให้ความสำคัญของปัจจัย A และ B เท่ากัน โดยพิจารณาทั้งปัจจัย A และ B เป็น Main Plot ทั้งคู่ และในการวิเคราะห์ ANOVA จะมี Mean Square Error (MSE) 3 ค่า ทำให้สามารถคำนวณ Coefficient of Variation (CV) ได้ 3 ค่า คือ

$$\text{CV (A)} = \frac{\sqrt{\text{MSE (A)}}}{\text{Grand Mean}}$$

$$\text{CV (B)} = \frac{\sqrt{\text{MSE (B)}}}{\text{Grand Mean}}$$

$$\text{CV (C)} = \frac{\sqrt{\text{MSE (C)}}}{\text{Grand Mean}}$$

ความแปรปรวนของการทดลองแบบ Strip Plot ซึ่งวางแผนการทดลองแบบ RCB สามารถแบ่งเป็นส่วนต่าง ๆ ได้ดังสมการ

$$\text{SST} = \text{SSA+SSE(A)+SSB+SSE(B)+SS(AB)+SSE(C)+SS(Block)} \quad \text{……………}(7.5)$$

7.3 One-Way ANOVA (Completely Randomized Design, CRD-การออกแบบการทดลองแบบสุ่มอย่างสมบูรณ์)

One-Way Classification Fixed Effect Model มีปัจจัยที่ต้องการทดสอบเพียงปัจจัยเดียว (ปัจจัยที่มีความสำคัญมากที่สุด) เช่น อุณหภูมิเป็นเพียงปัจจัยเดียวที่มีผลต่อการใช้น้ำของพืช ส่วนสภาพแวดล้อมอื่น ๆ นั้น เหมือนกัน (Homogeneity)

จึงทำการทดลองที่อุณหภูมิต่างกัน k ระดับ (k-treatment หรือ k-population) ระดับละ n ซ้ำ(โดยเลือกสุ่ม n ซ้ำอย่างสมบูรณ์) โดยสมมติว่า k-treatment (population) เป็น Independent and Normally Distributed (IND) มี Mean : $\mu_1, \mu_2, ..., \mu_k$ และ Common Variance = σ^2

7.3.1 ทดสอบ Hypothesis

H_o : $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$

H_1 : At least two μ_i are not equal

One-Way ANOVA

		k-Treatment (Population)					Total	Mean
		1	2	3	i	.....k		
	1	y_{11}	y_{21}	y_{31}	.......y_{i1}	.......y_{k1}	$T_{.1}$	$\bar{y}_{.1}$
	2	y_{12}	y_{22}	y_{32}	.......y_{i2}	......y_{k2}	$T_{.2}$	$\bar{y}_{.2}$
	3	y_{13}	y_{23}	y_{33}	.......y_{i3}	......y_{k3}	$T_{.3}$	$\bar{y}_{.3}$
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	j	y_{1j}	y_{2j}	y_{3j}	.......y_{ij}	......y_{kj}	$T_{.j}$	$\bar{y}_{.j}$
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	n	y_{1n}	y_{2n}	y_{3n}	.......y_{in}	......y_{kn}	$T_{.n}$	$\bar{y}_{.n}$
Total		$T_{1.}$	$T_{2.}$	$T_{3.}$	.......$T_{i.}$	......$T_{k.}$	$T_{..}$	
Mean $= T_{i.}/n$		$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$	.......$\bar{y}_{i.}$	......$\bar{y}_{k.}$		$\bar{y}_{..}=T_{..}/nk$
Population Mean		μ_1	μ_2	μ_3	μ_i	μ_k		$\mu=\sum_{i=1}^{k} \mu_i/k$

y_{ij} = j^{th} Observation from i^{th} treatment $\longrightarrow N(Y_{ij}\,;\,\mu_i, \sigma^2)$

y_{ij} = μ_i+E_{ij} (μ_i = mean of treatment i)

E_{ij} = $y_{ij}-\mu_i$ = Random Error/Sampling Error $\approx N(E_{ij}\,;\,0, \sigma^2)$

y_{ij} = μ_i+E_{ij}

= $\mu+(\mu_i-\mu)+E_{ij}$

y_{ij} = $\mu+\alpha_i+E_{ij}$ $(i = 1,2,.....,k\,;\, j = 1, 2,........,n)$ ……….(7.6)

μ = Grand Mean = $\left(\sum_{i=1}^{k} \mu_i\right)/k$

α_i = $\mu_i-\mu$ = Effect of Treatment i

Ho	:	$\alpha_i = 0$ (i = 1, 2,.............., k) (no treatment effect)
H_1	:	$\alpha_i \neq 0$ (i = 1, 2,.............., k) (treatment effect)

การทดสอบจะใช้วิธีการวิเคราะห์ความแปรปรวนของข้อมูล 2 แหล่ง คือ

(1) ความแปรปรวนเนื่องจาก Treatment หรือระหว่าง Treatment (Between Treatment) เป็นความแปรปรวนแบบ Systematic

(2) ความแปรปรวนภายใน Treatment (Within Treatment) เป็นความแปรปรวน Random

ความแปรปรวนรวม (Total Sum of Squares, SST)

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 \qquad \text{.....}(7.7)$$

ความแปรปรวนเนื่องจาก Treatment (Treatment Sum of Squares, SSA)

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 \qquad \text{.....}(7.8)$$

ความแปรปรวนภายในกลุ่ม (Within treatment) (Error Sum of Squares, SSE)

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 \qquad \text{.....}(7.9)$$

$$SST = SSA + SSE \qquad \text{.....}(7.10)$$

7.3.2 พิสูจน์ว่า SST = SSA+SSE

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{k}(\bar{y}_{i}-\bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}\{(y_{ij}-\bar{y}_{i.})+(\bar{y}_{i.}-\bar{y}_{..})\}^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})(\bar{y}_{i}-\bar{y}_{..}) + \sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2$$

เนื่องจากที่ i ใด ๆ $\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.}) = \sum_{j=1}^{n} y_{ij} - n\bar{y}_{i.} = 0$

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$$

$$SST = SSE + SSA \quad ....\text{Q.E.D.}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{k}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{T_{..}^2}{nk} \quad ....(7.11)$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 = \sum_{i=1}^{k}\frac{T_{i.}^2}{n} - \frac{T_{..}^2}{nk} \quad ....(7.12)$$

Treatment Mean Square, MSA

$$MSA = S_1^2 = \frac{SSA}{k-1} \quad ....(7.13)$$

หา $\mathbf{E(S^2_1) = E(\frac{SSA}{k-1})}$

$$y_{ij} = \mu+\alpha_i+E_{ij}$$

$$\bar{y}_{i.} = \sum_{j=1}^{n}\frac{(\mu+\alpha_i+E_{ij})}{n} = \mu+\alpha_i+\sum_{j=1}^{n}\frac{E_{ij}}{n}$$

$$= \mu+\alpha_i+\bar{E}_{i.}$$

$$\bar{y}.. = \sum_{i=1}^{k}\sum_{j=1}^{n}\frac{(\mu+\alpha_i+E_{ij})}{nk} = \mu+\frac{n}{nk}\sum_{i=1}^{k}\alpha_i+\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{E_{ij}}{nk}$$

$$= \mu+\frac{1}{k}\sum_{i=1}^{k}\cancelto{0}{(\mu_i-\mu)}+\bar{E}_{..}$$

$$= \mu+\bar{E}_{..}$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$$

$$= \sum_{i=1}^{k}(\cancel{\mu}+\alpha_i+\bar{E}_{i.}-\cancel{\mu}-\bar{E}_{..})^2$$

$$= n\sum_{i=1}^{k}\{\alpha_i+\bar{E}_i-\bar{E}_{..}\}^2$$

$$= n\sum_{i=1}^{k}(\alpha_i^2+\bar{E}_i^2-\bar{E}^2_{..}+2\,\alpha_i\,\bar{E}_{i.}-2\alpha_i\,\bar{E}_{..}-2\bar{E}_{i.}\,\bar{E}_{..})$$

$$E(SSA) = n\sum_{i=1}^{k}[E(\alpha_i^2)+E(\bar{E}_{i.})^2+E(\bar{E}_{..})^2+2E(\alpha_i\,\bar{E}_{i.})-2E(\alpha_i\,\bar{E}_{..})-2E(\bar{E}_{i.}\,\bar{E}_{..})] \quad ...(7.14)$$

(1) $\alpha_i = \mu_i-\mu$ ไม่ใช่ Random Variable เป็นค่าคงที่

(2) $E(\bar{E}_{i.})^2 = E(\sum_{j=1}^{n}\frac{E_{ij}}{n})^2$

$$= \frac{1}{n^2} E(E_{i1}+E_{i2}+\dots\dots\dots\dots+E_{in})^2$$

$$= \frac{1}{n^2} E[(E^2_{i1}+E^2_{i2}+\dots\dots\dots+E^2_{in})+2(E_{i1}.E_{i2}+\dots\dots+E_{i(n-1)}.E_{in})]$$

$$= \frac{1}{n^2}\left[\sum_{j=1}^{n} E(E_{ij})^2 + 2(E(\cancel{E_{i1}.E_{i2}})^{0})+\dots\dots\dots+E(\cancel{E_{i(n-1)}.E_{in}})^{0})\right]$$

E_{ij} = Independent Normally Distributed $\approx N(E_{ij}; 0, \sigma^2)$

$E(E_{i1}.E_{i2}) = E(E_{i1})E(E_{i2}) = 0$

$$E(\bar{E}_{i.})^2 = \frac{1}{n^2}\sum_{j=1}^{n} E(E_{ij}-0)^2 = \frac{1}{n^2}\sum_{j=1}^{n}\sigma^2 = \frac{\sigma^2}{n} \quad \dots\dots\dots\dots\dots\dots(7.15)$$

(3) $E(\bar{E}_{..})^2$

$$= E\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{E_{ij}}{nk}\right)^2$$

$$= \frac{1}{(nk)^2} E\left(\sum_{i=1}^{k}\sum_{j=1}^{n} E_{ij}^2\right)$$

$$= \frac{1}{(nk)^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{n} E(E_{ij}-0)^2 + \cancel{E(\text{Cross Product Terms})}^{0}\right]$$

$$= \frac{\sigma^2}{nk} \quad \dots\dots\dots(7.16)$$

(4) $E(\bar{E}_{i.}) = E\left(\sum_{j=1}^{n}\frac{E_{ij}}{n}\right) = \frac{1}{n}\sum_{j=1}^{n} E(E_{ij}) = 0 \quad \dots\dots\dots(7.17)$

(5) $E(\bar{E}_{..}) = E\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{E_{ij}}{nk}\right) = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} E(E_{ij}) = 0 \quad \dots(7.18)$

(6) $E(\bar{E}_{i.}\ \bar{E}_{..}) = E\left(\sum_{j=1}^{n}\frac{E_{ij}}{n}\right)\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{E_{ij}}{nk}\right)$

$$= \frac{1}{n^2k} E(E_{i1}+E_{i2}+\dots\dots\dots\dots+E_{in})(E_{11}+E_{12}+\dots\dots\dots\dots\dots+E_{kn})$$

$$= \frac{1}{n^2k} E(E^2_{i1}+E^2_{i2}+\dots\dots\dots\dots+E^2_{in}+ \text{Cross Product Terms})$$

$$= \frac{1}{n^2k}\left[\sum_{j=1}^{n} E(E_{ij}-0)^2 + E\,(\text{Cross Product Terms})\right]$$

0

$$= \frac{\sigma^2}{nk} \qquad \ldots\ldots\ldots(7.19)$$

$$E(SSA) = n\sum_{i=1}^{k}\left[\alpha_i^2+\frac{\sigma^2}{n}+\frac{\sigma^2}{nk}+0-0-2\frac{\sigma^2}{nk}\right]$$

$$E(SSA) = n\sum_{i=1}^{k}\alpha_i^2 + k\sigma^2-\sigma^2$$

$$= (k-1)\,\sigma^2+n\sum_{i=1}^{k}\alpha_i^2$$

$$E\left(\frac{SSA}{k-1}\right) = \sigma^2+\frac{n\sum_{i=1}^{k}\alpha_i^2}{k-1} \qquad \ldots\ldots\ldots(7.20)$$

ถ้า H_o : $\alpha_i = 0$ เป็นจริง

$$E(S_1^2) = E\left(\frac{SSA}{k-1}\right) = \sigma^2 \qquad \ldots\ldots\ldots(7.21)$$

แสดงว่า $S_1^2 = \frac{SSA}{k-1}$ เป็น Unbiased Estimate ของ σ^2

Error Mean Square, MSE

$$MSE = S^2 = \frac{SSE}{k(n-1)} \qquad \ldots\ldots\ldots(7.22)$$

S^2 เป็น Unbiased Estimate ของ σ^2 ไม่ว่า Ho จะเป็นจริงหรือไม่

$f = \frac{S_1^2}{S^2}$ เป็น Random Variable

F-distribution with (k-1) และ k(n-1) degree of freedom

If $f > f_{\alpha,[k-1,\,k(n-1)]}$ Reject Ho : $\alpha_i \neq 0$ (one tail test)

$$SSA = \sum_{i=1}^{k}\frac{T_{i.}^2}{n}-\frac{T_{..}^2}{nk} \qquad \ldots\ldots\ldots(7.23)$$

$$= \sum_{i=1}^{k} n\,\bar{y}_{i.}^2-\frac{T_{..}^2}{nk}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2-\frac{T_{..}^2}{nk} \qquad \ldots\ldots\ldots(7.24)$$

One Way – ANOVA

		k-Treatment					Total	Mean
		1	2	3	i	.....k		
	1	y_{11}	y_{21}	y_{31}	.......y_{i1}	.......y_{k1}	$T_{.1}$	$\bar{y}_{.1}$
	2	y_{12}	y_{22}	y_{32}	.......y_{i2}	......y_{k2}	$T_{.2}$	$\bar{y}_{.2}$
	3	y_{13}	y_{23}	y_{33}	.......y_{i3}	......y_{k3}	$T_{.3}$	$\bar{y}_{.3}$
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	j	y_{1j}	y_{2j}	y_{3j}	.......y_{ij}	......y_{kj}	$T_{.j}$	$\bar{y}_{.j}$
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	n	y_{1n}	y_{2n}	y_{3n}	.......y_{in}	......y_{kn}	$T_{.n}$	$\bar{y}_{.n}$
Total		$T_{1.}$	$T_{2.}$	$T_{3.}$	.......$T_{i.}$	......$T_{k.}$	$T_{..}$	
Mean		$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$	.......$\bar{y}_{i.}$	......$\bar{y}_{k.}$		$\bar{y}_{..}$
$\sum \frac{T_{i.}^2}{n}$		$\frac{T_{1.}^2}{n}$	$\frac{T_{2.}^2}{n}$	$\frac{T_{3.}^2}{n}$	$\frac{T_{i.}^2}{n}$	$\frac{T_{k.}^2}{n}$		

7.3.3 ANOVA for One Way Classification (CRD)

Source of Variation	Sum of Squares	D.F	Mean Square	Computed f
Treatment	$\sum_{i=1}^{k} \frac{T_{i.}^2}{n} - \frac{T_{..}^2}{nk}$	k-1	$S_1^2 = \frac{SSA}{k-1}$	$\frac{S_1^2}{S^2}$
Error	SST-SSA	k(n-1)	$S^2 = \frac{SSE}{k(n-1)}$	
Total	$\sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^2 - \frac{T_{..}^2}{nk}$	nk-1		

ตัวอย่างที่ 7.1 จงทดสอบว่าเส้นทางจากบ้านไปมหาวิทยาลัย 4 เส้นทาง ใช้เวลาเฉลี่ยเท่ากันหรือไม่ โดยการสุ่มตัวอย่าง เส้นทางละ 5 ตัวอย่าง ดังตาราง

j	4 Treatments			
	เส้นทางที่ 1	เส้นทางที่ 2	เส้นทางที่ 3	เส้นทางที่ 4
1	22	25	26	26
2	26	27	29	28
3	25	28	33	27
4	25	26	30	30
5	31	29	33	30
Ti.	129	135	151	141
T..	556			

$$\sum_{i=1}^{k} T_{i.}^2 = 77,548$$

$$\sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij}^2 = 15,610$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^2 - \frac{T_{..}^2}{nk}$$

$$= 15,610 - \frac{(556)^2}{5 \times 4}$$

$$= 153.2$$

$$SSA = \sum_{i=1}^{k} \frac{T_{i.}^2}{n} - \frac{T_{..}^2}{nk}$$

$$= \frac{77,548}{5} - \frac{(556)^2}{5 \times 4}$$

$$= 52.8$$

$$SSE = 153.2 - 52.8$$

$$= 100.4$$

ทดสอบ H_o : $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1 : μ ไม่เท่ากันอย่างน้อย 1 คู่

$\alpha = 0.05$

SOURCE	SS	SF	MS.	Computed f
Treatment	52.8	4-1	$\frac{52.8}{3} = 17.6$	$\frac{17.6}{6.28} = 2.80$
Error	100.4	4(5-1)	$\frac{100.4}{16} = 6.28$	
Total	153.2	5 x 4-1		

$f_{0.05\,(3,\,16)} = 3.24$

Computed f $<$ $f_{0.05\,(3,\,16)}$

Accept Ho หรือ $\mu_1 = \mu_2 = \mu_3 = \mu_4$ ที่ $\alpha = 0.05$

(μ_i ไม่ต่างกันอย่างมีนัยสำคัญทางสถิติ)

7.4 ANOVA กรณี Unequal Sample Size

กรณีที่มี k Treatment แต่ละ Treatment มีขนาด $n_1, n_2, \ldots\ldots, n_k$ และ $N = n_1+n_2+\ldots\ldots+n_k$

(จาก $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}\frac{x_i}{n} = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}\bar{x}^2$)

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{ni}(y_{ij} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{ni}(y_{ij}^2 - 2y_{ij}\bar{y}_{..} + \bar{y}_{..}^2)$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{ni} y_{ij}^2 - 2\bar{y}_{..}\sum_{i=1}^{k}\sum_{j=1}^{ni} y_{ij} + N\bar{y}_{..}^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{ni} y_{ij}^2 - N\bar{y}_{..}^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7.25)$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{ni} y_{ij}^2 - \frac{T_{..}^2}{N} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7.26)$$

$$SSA = \sum_{i=1}^{k}\sum_{j=1}^{ni}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{k} n_i(\bar{y}_{i.}^2 - 2\bar{y}_{i.}\bar{y}_{..} + \bar{y}_{..}^2)$$

$$= \sum_{i=1}^{k} n_i\bar{y}_{i.}^2 - 2\bar{y}_{..}\sum_{i=1}^{k} n_i\bar{y}_{i.} + N\bar{y}_{..}^2$$

$$= \sum_{i=1}^{k} n_i \left(\frac{T_{i.}}{n_i}\right)^2 - 2\bar{y}_{..} \sum_{i=1}^{k} n_i \left(\frac{T_{i.}}{n_i}\right) + N\left(\frac{T_{..}}{N}\right)^2$$

$$= \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - 2\frac{T_{..}}{N} T_{..} + \frac{T_{..}^2}{N}$$

$$= \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N} \qquad \text{.....................................(7.27)}$$

$$SSE = SST - SSA \qquad \text{.....................................(7.28)}$$

Degree of Freedom

D.F. of SST = N-1

SSA = k-1

SSE = (N-1)-(k-1) = N-k ....(7.29)

ตัวอย่างที่ 7.2 ในการทดสอบว่าค่าเฉลี่ยของ % กำมะถัน ในถ่านหินของแหล่งสำรวจ 5 แหล่งเท่ากันหรือไม่ ได้ทำการสุ่มตัวอย่างจากถ่านหิน 5 แหล่ง เพื่อวิเคราะห์หา % ของกำมะถันปรากฏผลดังนี้

i / j	1	2	3	4	5
1	1.51	1.69	1.56	1.30	0.73
2	1.92	0.64	1.22	0.75	0.80
3	1.08	0.90	1.32	1.26	0.90
4	2.04	1.41	1.39	0.69	1.24
5	2.14	1.01	1.33	0.62	0.82
6	1.76	0.84	1.54	0.90	0.72
7	1.17	1.28	1.04	1.20	0.57
8		1.59	2.25	0.32	1.18
9			1.49		0.54
10					1.30
Ti.	11.62	9.36	13.14	7.04	8.80
T..	49.96				

N = 7+8+9+8+10 = 42

k = 5

H_o = $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_1 = อย่างน้อยค่าเฉลี่ยจาก 2 แหล่งไม่เท่ากัน

α = 0.01

$f_{0.01,(4,37)}$ = 3.83

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{ni} y_{ij}^2 - \frac{T_{..}^2}{N}$$

$$= 67.861 - \frac{(49.96)^2}{42} = 8.432$$

$$SSA = \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N}$$

$$= \frac{(11.62)^2}{7} + \frac{(9.36)^2}{8} + \frac{(13.14)^2}{9} + \frac{(7.04)^2}{8} + \frac{(8.8)^2}{10} - \frac{(49.96)^2}{42}$$

= 3.935

SSE = SST-SSA

= 8.432-3.935 = 4.497

$$MSA = S_1^2 = \frac{SSA}{k-1} = \frac{3.935}{4} = 0.984$$

$$MSE = S^2 = \frac{SSE}{N-k} = \frac{4.497}{37} = 0.122$$

ANOVA

SOURCE	SS.	DF.	MS.	Computed f
TREATMETN	3.935	4	0.984	0.984/0.122
ERROR	4.497	37	0.122	= 8.066
TOTAL	8.432	41		

$f > f_{0.01,(4,37)}$

Reject H_o

7.5 การเปรียบเทียบ Variances

ในการทดสอบ Mean ด้วย F-TEST ตามที่กล่าวมาแล้ว

H_o : $\mu_1=\mu_2=\mu_3=\ldots\ldots=\mu_k$

H_1 : μ อย่างน้อย 1 คู่ไม่เท่ากัน

มีสมมติฐานว่า

$\sigma_1=\sigma_2=\ldots\ldots=\sigma_k = \sigma$

ถ้า n_i $(i = 1,2,\ldots,k)$ เท่ากัน จะเกิด Error น้อยที่สุด เพราะถึงแม้ว่า σ_i ไม่เท่ากัน ก็ต่างกันไม่มาก

ถ้า n_i ไม่เท่ากัน และ σ_i ไม่เท่ากัน ค่า σ_i บางค่าอาจทำให้การคำนวณ F ผิดพลาด และมีผลทำให้การทดสอบผิดพลาดได้ ดังนั้น กรณีที่ n_i ไม่เท่ากันจึงต้องมีการทดสอบว่า σ_i เท่ากันหรือไม่

H_o : $\sigma_1 = \sigma_2 = \ldots\ldots = \sigma_k$

Test of Equality of Several Variance

$B \sim \chi^2_{\alpha,\, k-1}$

สามารถทดสอบได้ 2 วิธี

วิธีที่ 1 : ใช้ χ^2-Test

1. คำนวณ S_1^2 , S_2^2, ..., S_k^2

2. หา $MSE = Sp^2 = \sum_{i=1}^{k} \frac{(n_i-1)S_i^2}{(N-k)}$ ……………………………..(7.30)

3. $b = 2.3026\,\frac{Q}{h}$ ……………………..(7.31)

เมื่อ $Q = (N-k)\log Sp^2 - \sum_{i=1}^{k}(n_i-1)\log S_i^2$ ………………………..(7.32)

ถ้า $S_1 \approx S_2 \approx \ldots\ldots \approx S_k$; $Q \rightarrow 0$

$h = 1+\frac{1}{3(k-1)}\left\{\left(\sum_{i=1}^{k}\frac{1}{(n_i-1)}\right)-\left(\frac{1}{N-k}\right)\right\}$ ……………..(7.33)

ถ้า S_i ต่างกันมาก Q จะมีค่ามาก b จะมากตาม

$b > \chi^2_{\alpha,\,(k-1)} \rightarrow$ Reject Ho

วิธีที่ 2 : ใช้ B-Test

1. คำนวณ b

$$= \frac{[(S_1^2)^{n_1 - 1}(S_2^2)^{n_2 - 1} \ldots\ldots\ldots(S_k^2)^{n_k - 1}]^{\frac{1}{N-k}}}{Sp^2} \quad \ldots\ldots\ldots\ldots\ldots(7.34)$$

2. ทดสอบ

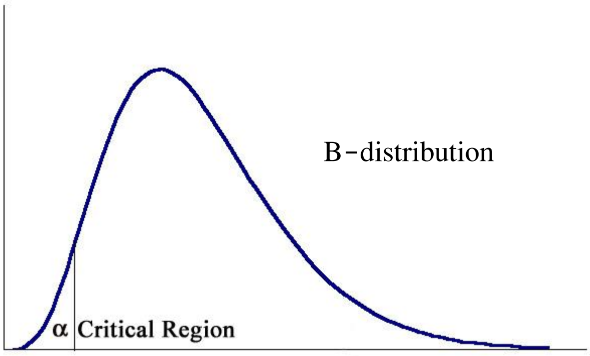
2.1 **กรณีที่ 1** : $n_1 = n_2 = \ldots\ldots\ldots\ldots\ldots = n_k$

ถ้า $b < b_k(\alpha;n)$ (ตารางที่ A9 ของภาคผนวก) Reject Ho

2.2 **กรณีที่ 2** : n_i ไม่เท่ากัน

Reject Ho : ถ้า $b < b_k(\alpha;n_1, n_2, \ldots\ldots\ldots, n_k)$

$$b_k(\alpha;n_1, n_2, \ldots, n_k) = \frac{[n_1 b_k(\alpha;\ n_1) + n_2 b_k(\alpha;\ n_2) + \ldots\ldots + n_k b_k(\alpha;\ n_k)]}{N} \quad \ldots\ldots(7.35)$$



ตัวอย่างที่ 7.3 จากตัวอย่าง 11.2 กรณี n_i ไม่เท่ากัน จึงต้องทดสอบ σ_i ของแหล่งถ่านหินทั้ง 5 ใช้ α = 0.01

H_o : $\sigma_1^2 = \sigma_2^2 = \ldots\ldots\ldots\ldots\ldots = \sigma_k^2$

H_1 : σ_i^2 อย่างน้อย 1 คู่ไม่เท่ากัน

$\chi^2_{0.01, 4}$ = 13.3

ทดสอบด้วย χ^2 Test

MSE $= S_p^2 = \sum_{i=1}^{k} \frac{(ni-1)S_i^2}{N-k}$

$= \frac{4.504}{42-5} = 0.1217$

Q $= (N\text{-}k)\log S_p^2 - \sum_{i=1}^{k}(n_i - 1)\log S_i^2$

$= (42\text{-}7)\log(0.1217) - (-34.498)$

$= -33.844 + 34.498 = 0.6538$

h $= 1 + \frac{1}{3 \times 4}\left\{\left(\sum_{i=1}^{k}\frac{1}{n_i - 1}\right) - \frac{1}{N-k}\right\}$

$= 1 + \frac{1}{3 \times 4}\left\{\frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{7} + \frac{1}{9} - \frac{1}{37}\right\}$

$= 1.055$

b $= 2.3026\,\frac{Q}{h} = 2.3026 \times \frac{0.6538}{1.055}$

$= 1.427$

$b\ (= 1.427) < \chi^2_{0.01,4}\ (=13.3)$

Accept Ho : $\sigma_1^2 = \sigma_2^2 = \ldots\ldots\ldots\ldots\ldots = \sigma_k^2$

ที่ α = 0.01 (ไม่สามารถสรุปได้ว่า σ_i^2 ต่างกันอย่างมีนัยสำคัญ)

ทดสอบด้วย B Test

b $= \frac{\left[(S_1^2)^{n_1-1}(S_2^2)^{n_2-1}\ldots\ldots\ldots(S_k^2)^{n_k-1}\right]^{\frac{1}{N-k}}}{S_p^2}$

b $= \frac{\left[(0.175)^6(0.144)^7\ldots\ldots\ldots(0.074)^9\right]^{\frac{1}{42-5}}}{0.1217}$

$= 0.9603$

$b_k\ (\alpha\,; n_1, n_2, \ldots\ldots\ldots, n_5)$

$= \frac{n_1 b_k(\alpha; n_1) + n_2 b_k(\alpha; n_2) + \ldots + n_k b_k(\alpha; n_k)}{N}$

จากตาราง Bartlett's Statistics (α = 0.01)

b_5 (0.01;7) = 0.6248

b_5 (0.01;8) = 0.6704

b_5 (0.01;9) = 0.7062

b_5 (0.01;10) = 0.7352

b_5 (0.01 ; 7, 8, 9, 8, 10)

$$= \frac{7 \text{ x } 0.6248 + 8 \text{ x } 0.6704 + 9 \text{ x } 0.7062 + 8 \text{ x } 0.6704 + 10 \text{ x } 0.7352}{42}$$

= 0.6866

b(0.9603) > b_5(0.01 ; 7, 8, 9, 8, 10)

Accept H_o

ตัวอย่างที่ 7.4 จงทดสอบ σ^2 ของค่าในตารางด้วย B Test โดยใช้ค่า Bartlett วิกฤต กำหนดให้ α = 0.01

Treatments (k=4)

1		2	3	4
49.20	97.50	97.07	62.10	110.60
44.54	105.0	73.40	94.95	57.10
45.80	58.05	68.50	142.50	117.60
95.84	86.60	91.85	53.00	77.71
30.10	58.35	106.60	175.00	150.00
36.50	72.80	0.57	79.50	82.90
82.30	116.70	0.79	29.50	111.50
87.85	45.15	0.77	78.40	-
105.00	70.35	0.81	127.50	-
95.22	77.40	-	-	-
20		9	9	7

N = 45

H_o : $\sigma^2_1 = \sigma^2_2 = \ldots\ldots\ldots\ldots\ldots\ldots = \sigma^2_4$

H_1 : อย่างน้อย σ^2_1 1 คู่ ไม่เท่ากัน

หา b วิกฤต : b_k (α ; n_1, n_2, n_3, n_4)

b_4 (0.01 ; 20, 9, 9, 7)

$$= \frac{20 \text{ x } 0.8586 + 9 \text{ x } 0.6892 + 9 \text{ x } 0.6892 + 7\text{x}0.6042}{45}$$

b_4 (0.01 ; 20, 9, 9, 7) = 0.7513

คำนวณค่า b

$$S_1^2 = 662.862$$

$$S_2^2 = 2{,}219.781$$

$$S_3^2 = 2{,}168.032$$

$$S_4^2 = 946.032$$

$$S_p^2 = \sum_{i=1}^{k} \frac{(n_i - 1) S_i^2}{N - k}$$

$$= \frac{19\,(662.862) + 8(2{,}219.781) + 8(2{,}168.434) + 6(946.032)}{45 - 4}$$

$$= 1{,}301.861$$

$$b = \frac{\left\{(662.862)^{19} (2219.781)^{8} (2168.434)^{8} (946.032)^{6}\right\}^{1/45-4}}{1301.861}$$

$$= 0.8557$$

b (~0.8557) > b_4(0.01 ; 20, 9, 9, 7) (~ 0.7513)

Accept H_o : $\sigma^2_1 = \sigma^2_2 = \ldots\ldots\ldots\ldots\ldots\ldots = \sigma^2_4$

Cochran'S Test (วิธีทดสอบแบบง่าย ๆ)

H_o : $\sigma^2_1 = \sigma^2_2 = \ldots\ldots\ldots\ldots\ldots\ldots = \sigma^2_k$

ถ้า n_i เท่ากัน

$$g = \frac{\text{ค่า } S_i^2 \text{ ที่มีค่ามากที่สุด}}{\sum_{i=1}^{k} S_i^2}$$

ถ้า $g > g_k\,(\alpha; n) \rightarrow$ Reject H_o

ค่า $g_k\,(\alpha; n)$ ดูได้จากตาราง A10 ของภาคผนวก

ตัวอย่างที่ 7.5 จากผลการทดสอบ การดูดซึมความชื้นของคอนกรีต ที่มีส่วนผสมต่างกัน 5 สูตร ดังตาราง จงทดสอบว่า σ^2_i เท่ากันที่ $\alpha = 0.05$

% การดูดซึมความชื้นของคอนกรีต (% โดยน้ำหนัก)					
	1	2	3	4	5
	551	595	639	417	563
	457	580	615	449	631
	450	508	511	517	522
	731	583	573	438	613
	499	633	648	415	656
	632	517	677	555	679
s_i^2	12,134	2,303	3,594	3,319	3,455

H_o : $\sigma^2_1 = \sigma^2_2 = \ldots\ldots\ldots\ldots = \sigma^2_5$

H_1 : σ^2_i ไม่เท่ากันทุกค่า

$$g = \frac{MAX(S_i^2)}{\sum_{i=1}^{k} S_i^2} = \frac{12{,}134}{24{,}805} = 0.4892$$

$g_5\,(0.05\,;\,6) = 0.5065$

$g < g_5(0.05;6)$

Accept H_o

7.6 การออกแบบการทดลองแบบสุ่มอย่างสมบูรณ์ในบล๊อค

(Randomized Complete Block Design)

ในการวิเคราะห์ ANOVA กรณีออกแบบการทดลองแบบสุ่มอย่างสมบูรณ์ในแต่ละบล๊อค (RCB) สาเหตุของความแปรปรวน อาจเกิดจาก

(1) วิธีปฏิบัติ (Treatment)

ซึ่งคาดว่าจะมีอิทธิพลต่อข้อมูลมากที่สุด

(2) วิธีดำเนินการทดลอง (Block)

เช่น ต้องการศึกษาว่าปริมาณน้ำชลประทานมีผลต่อผลผลิตพืช โดยการทดลองปลูกพืชแล้วให้น้ำด้วยอัตราต่าง ๆ กัน k ระดับ (k-treatments) ถ้าพิจารณาว่าสภาพแวดล้อมของการทดลองเช่นดิน ไม่ต่างกัน (Homogeneity) จะออกแบบทดลองแบบ CRD โดยทำการทดลอง n ซ้ำ (n samples) ในแต่ละ treatment แบบสุ่มอย่างสมบูรณ์ ตามที่กล่าวมาแล้วในข้อ 7.2

แต่ถ้าพิจารณาว่าดินในแปลงทดลองมีความแตกต่างกัน (Heterogeneity) จะแบ่งแปลงทดลองตามสภาพดินออกเป็น b กลุ่ม (Blocks) แต่ละกลุ่มทดลอง k treatment โดยสุ่มอย่างสมบูรณ์ในแต่ละกลุ่ม ซึ่งจะมีผลทำให้ Degree of Freedom ลดลง จาก k(n-1) เป็น (k-1)(n-1)

สมมติให้

treatment = k = 4

block = b = 4 หรือ sample = n = 4

$\therefore$ แปลงทดลอง = 4 x 4 = 16 แปลง

CRD

t4(1)	t3(2)	t1(1)	t4(3)
t1(4)	t1(2)	t3(3)	t3(1)
t2(2)	t2(1)	t4(2)	t1(3)
t4(4)	t2(3)	t2(4)	t3(4)

DF of Error = k(n-1) = 4(4-1) = 12

RCB

B1	t1	t4	t2	t3
B2	t4	t1	t3	t2
B3	t1	t3	t2	t4
B4	t3	t1	t4	t2

DF of Error = (k-1)(n-1) = (4-1)(4-1) = 9

หรือ ในการทดลองว่าเวลาเฉลี่ยในการผลิตผลิตภัณฑ์ชนิดหนึ่ง จากเครื่องจักร 4 เครื่อง เท่ากันหรือไม่ ปัจจัยที่คาดว่ามีอิทธิพลต่อเวลาเฉลี่ยมากที่สุด คือ ความแตกต่างของเครื่องจักรทั้ง 4 (4 treatments) แต่ปัจจัยที่อาจมีผลต่อเวลาเฉลี่ยอาจได้แก่ วิธีดำเนินการ (Block) เช่น พนักงานหรือเวลา (กะกลางวัน/กะกลางคืน) ในการผลิต หรืออื่น ๆ ซึ่งไม่ Homogeneity ดังนั้นจึงต้องทำการทดสอบแบบ RCB เพื่อขจัดอิทธิพลของวิธีดำเนินการ (Block)

7.6.1 สมมติฐานในการทดลอง

1. Equal Treatment Means

H_o : $\mu_{1.} = \mu_{2.} = \ldots\ldots\ldots\ldots = \mu_{k.} = \mu$ (No treatment effect)

$\mu_{i.}$ = mean of treatment i

2. Equal Block Means

H'_o : $\mu_{.1} = \mu_{.2} = ... = \mu_{.b} = \mu$ (No block effect)

$\mu_{.j}$ = mean of block j

X_{ij} = เวลาเฉลี่ยในการผลิตผลิตภัณฑ์จากเครื่องจักร (Treatment) i และวิธีดำเนินงาน (Block) j

เมื่อ i = 1, 2,................, k

j = 1, 2,................, b

N = kb

RCB-Design

วิธีดำเนินการ (Block)	วิธีปฏิบัติหรือระบบ Treatment							รวม	เฉลี่ย
	1	2	3	.	i	.	k		
1	x_{11}	x_{21}	x_{31}	.	x_{i1}	.	x_{k1}	$T_{.1}$	$\bar{x}_{.1}$
2	x_{12}	x_{22}	x_{32}	.	x_{i2}	.	x_{k2}	$T_{.2}$	$\bar{x}_{.2}$
3	x_{13}	x_{23}	x_{33}	.	x_{i3}	.	x_{k3}	$T_{.3}$	$\bar{x}_{.3}$
.	.	.	.	.	.	.	.	.	.
j	x_{1j}	x_{2j}	x_{3j}	.	x_{ij}	.	x_{kj}	$T_{.j}$	$\bar{x}_{.j}$
.	.	.	.	.	.	.	.	.	.
b	x_{1b}	x_{2b}	x_{3b}	.	x_{ib}	.	x_{kb}	$T_{.b}$	$\bar{x}_{.b}$
รวม	$T_{1.}$	$T_{2.}$	$T_{3.}$	.	$T_{i.}$	.	$T_{k.}$	$T_{..}$	
เฉลี่ย	$\bar{x}_{1.}$	$\bar{x}_{2.}$	$\bar{x}_{3.}$	.	$\bar{x}_{i.}$	.	$\bar{x}_{k.}$		$\bar{x}_{..}$

$T_{i.}$ = Sum of Treatment (i)

$$= \sum_{j=1}^{b} x_{ij} \qquad \text{..................(7.36)}$$

$$\bar{x}_{i.} = \frac{T_{i.}}{b} \qquad \text{..................(7.37)}$$

$T_{.j}$ = Sum of Block j

$$= \sum_{i=1}^{k} x_{ij}$$

$$\bar{x}_{.j} = \frac{T_{.j}}{k} \qquad \text{..................(7.38)}$$

$$T_{..} = \sum_{i=1}^{k}\sum_{j=1}^{b} x_{ij}$$

$$= \sum_{i=1}^{k} T_{i.} = \sum_{j=1}^{b} T_{.j} \qquad \text{………………(7.39)}$$

$$\bar{x}_{..} = \frac{T_{..}}{N} \qquad \text{………………(7.40)}$$

ถ้าหากมีความแตกต่างระหว่างวิธีปฏิบัติหรือระบบ (Treatment) หรือวิธีดำเนินการ (Block) จะต้องทดสอบว่า ความแตกต่างมากพอที่จะสรุปว่ามีผลทำให้ค่าเฉลี่ยของ Treatment หรือ Blcok แตกต่างจากค่าเฉลี่ยรวม (Grand Mean, μ) หรือไม่ โดยทำการทดสอบสมมติฐานดังนี้

H_o = $\mu_{1.} = \mu_{2.} = \dots\dots = \mu_{k.} = \mu$

H_o = $\mu_{.1} = \mu_{.2} = \dots\dots = \mu_{.b} = \mu$

μ_{ij} = ค่าเฉลี่ยของ Treatment i และ Block j

$\mu_{i.}$ = ค่าเฉลี่ยของวิธีปฏิบัติ (treatment) i = $\sum_{j=1}^{b}\frac{\mu_{ij}}{b}$ ………(7.41)

$\mu_{.j}$ = ค่าเฉลี่ยของวิธีดำเนินการ (block) j = $\sum_{i=1}^{k}\frac{\mu_{ij}}{k}$ ……….(7.42)

μ = Grand Mean

α_i = $\mu_{i.} - \mu$ = Effect of Treatment i ………………..………(7.43)

β_j = $\mu_{.j} - \mu$ = Effect of Block j ………………(7.44)

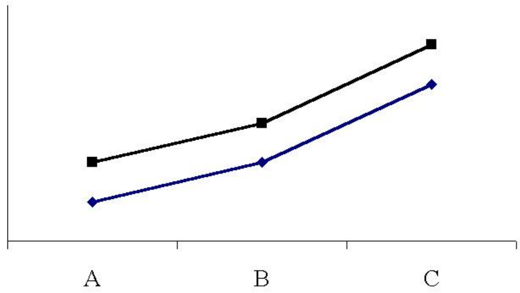
E_{ij} = $x_{ij} - \mu_{ij}$ = Unexplained Source of Effect (Random/Sampling Error)

x_{ij} = $\mu_{ij} + E_{ij}$ ………………(7.45)

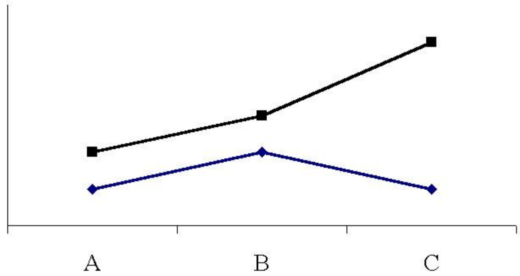
μ_{ij} = $\mu + \alpha_i + \beta_i$ ………………(7.46)

7.6.2 สมมติฐานในการทดสอบ

1. x_{ij} เป็น observation จำนวน kb ค่า ซึ่งเป็น Independent และ Randomly Sampling จาก kb populations ซึ่งไม่ทราบค่าเฉลี่ย
2. ทั้ง kb populations มีการแจกแจงแบบ Normal ; $N(x_{ij} ; \mu_{ij}, \sigma^2)$
3. ทั้ง kb population มี Variance = σ^2
4. ผลกระทบของ Treatment และ Block เป็น Additive แต่ไม่มีผลร่วม (Interaction) ดูรูปที่ 7.4



(a) Additive



(b) Interaction

รูปที่ 7.4 อิทธิพลของ Treatment และ Block แบบ (a) additive (b) interaction

$$x_{ij} = \mu+\alpha_i+\beta_j+E_{ij} \qquad \text{.............(7.47)}$$

$$x_{ij} - \mu = \alpha_i+\beta_i+E_{ij}$$

$$= (\mu_{i.}- \mu)+(\mu_{.j}-\mu)+E_{ij}$$

$$x_{ij}-\mu = (\mu_{i.}- \mu)+(\mu_{.j}-\mu)+\{x_{ij}-[(\mu+(\mu_{i.}- \mu)+(\mu_{.j}-\mu)]\} \qquad \text{........................(7.48)}$$

$$x_{ij}-\bar{x}_{..} = (\bar{x}_{i.}- \bar{x}_{..})+(\bar{x}_{.j}-\bar{x}_{..})+\{x_{ij}-[\bar{x}_{..}+(\bar{x}_{i.}- \bar{x}_{..})+(\bar{x}_{.j}-\bar{x}_{..})]\} \qquad \text{........................(7.49)}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{b}(x_{ij}-\bar{x}_{..})^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{b}\{(\bar{x}_{i.}-\bar{x}_{..})+(\bar{x}_{.j}-\bar{x}_{..})+\{x_{ij}-[\bar{x}_{..}+(\bar{x}_{i.}-\bar{x}_{..})+(\bar{x}_{.j}-\bar{x}_{..})]\}\}^2$$

$$= \sum_{i=1}^{k} b(\bar{x}_{i.}-\bar{x}_{..})^2+\sum_{j=1}^{b} k(\bar{x}_{.j}-\bar{x}_{..})^2+\sum_{i=1}^{k}\sum_{j=1}^{b}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2 \text{ ..(Additive)}$$

$$\text{...............(7.50)}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{b}(x_{ij}-\bar{x}_{..})^2 \qquad \text{..................(7.51)}$$

Treatment Sum of Square [SSA]

$$SSA = \sum_{i=1}^{k} b(\bar{x}_{i.} - \bar{x}_{..})^2 \qquad \text{.................(7.52)}$$

Block Sum of Square [SSB]

$$SSB = \sum_{j=1}^{b} k(\bar{x}_{.j} - \bar{x}_{..})^2 \qquad \text{.................(7.53)}$$

Error Sum of Square/Residual [SSE]

$$SSE = SST\text{-}SSA\text{-}SSB \qquad \text{.................(7.54)}$$

การทดสอบว่าไม่มีความแปรปรวน หรือความแตกต่างเนื่องจาก Treatment

$$H_o : \mu_{1.} = \mu_{2.} = \ldots\ldots = \mu_{k.} = \mu$$

$$\mu = \frac{\sum_{i=1}^{k}\sum_{j=1}^{b}\mu_{ij}}{kb}$$

$$= \sum_{i=1}^{k} \mu_{i.} / k \qquad \text{.................(7.55)}$$

$$\alpha_i = \mu_{i.} - \mu$$

$$H_o = \alpha_1 = \alpha_2 = \ldots\ldots = \alpha_k = 0$$

การทดสอบว่าไม่มีความแปรปรวน หรือความแตกต่างเนื่องจาก Block

$$H_o' = \mu_{.1} = \mu_{.2} = \ldots\ldots = \mu_{.b} = \mu$$

หรือ $H_o' = \beta_1 = \beta_2 = \ldots\ldots = \beta_b = 0$

Take Expected Value

DF ของ SSA $= k-1$

$$MSA = \frac{SSA}{k-1}$$

$$E(MSA) = E\left(\frac{SSA}{k-1}\right)$$

$$= \sigma^2 + \frac{b\sum_{i=1}^{k}\alpha_i^2}{k-1} \qquad \text{.......................(7.56)}$$

DF ของ SSB $= b-1$

$$MSB = \frac{SSB}{b-1}$$

$$E(MSB) = E\left(\frac{SSB}{b-1}\right)$$

$$= \sigma^2 + \frac{k\sum_{j=1}^{b}\beta_j^2}{b-1} \qquad \text{..................(7.57)}$$

DF ของ Total = kb-1

DF ของ SSE = (kb-1)-(k-1)-(b-1)

= (k-1)(b-1) ....(7.58)

(เหมือน DF ของ χ^2 ในการทดสอบ Independence)

$$\text{MSE} = \frac{\text{SSE}}{(k-1)(b-1)}$$

$$E(\text{MSE}) = E\left(\frac{\text{SSE}}{(k-1)(b-1)}\right) = \sigma^2 \qquad \text{..................(7.59)}$$

ทดสอบ Treatment Effect

$$f_1 = \frac{\text{MSA}}{\text{MSE}} \qquad \text{..................(7.60)}$$

ถ้า $H_o : \alpha_i = 0$ เป็นจริง

$E(\text{MSA}) = E(\text{SSE}) = \sigma^2$

f_1 = 1 (Accept H_o)

ถ้า $\alpha_i \neq 0$: $f_1 > 1$

ถ้า $f_1 > f_{\alpha, [(k-1), (k-1)(b-1)]}$ (Reject H_o)

ทดสอบ Block Effect

$$f_2 = \frac{\text{MSB}}{\text{MSE}} \qquad \text{..................(7.61)}$$

ถ้า $H_o' \rightarrow$: $\beta_j = 0$ เป็นจริง

$E(\text{MSB}) = E(\text{SSE}) = \sigma^2$ ....(7.62)

f_2 = 1 (Accept H_o')

ถ้า $f_2 > f_{\alpha, [(b-1), (k-1)(b-1)]}$ (Reject H_o')

ANOVA กรณีสุ่มอย่างสมบูรณ์ในแต่ละบล๊อค (Randomized Complete Block)

SOURCES OF VARIATION	DF	SS	MS	F
(A) Treatment	k-1	$\sum_{i=1}^{k} \frac{T_{i.}^2}{b} - \frac{T_{..}^2}{N}$	$\frac{SSA}{k-1}$	$f_1 = \frac{MSA}{MSE}$
(B) Block	b-1	$\sum_{j=1}^{b} \frac{T_{.j}^2}{k} - \frac{T_{..}^2}{N}$	$\frac{SSB}{b-1}$	$f_2 = \frac{MSB}{MSE}$
(E) Error	(k-1)(b-1)	SST-SSA-SSB	$\frac{SSE}{(k-1)(b-1)}$	
(T) Total	kb-1	$\sum_{i=1}^{k} \sum_{j=1}^{b} x_{ij}^2 - \frac{T_{..}^2}{N}$		

ตัวอย่างที่ 7.6 จงทดสอบการสึกหรอของยางรถยนต์โดยเฉลี่ย หลังจากใช้งานแล้ว 20,000 กม. จากผลการสุ่มตัวอย่างยางรถยนต์ 4 ชนิด ซึ่งทดสอบการใช้งานในเส้นทางแตกต่างกัน 5 เส้นทาง ได้ผลดังตารางทดสอบโดยใช้ α = 0.01

เส้นทาง	การสึกหรอของยางรถ				รวม	ค่าเฉลี่ย
	1	2	3	4	T.j	
1	9.1	17.1	20.8	11.4	58.8	14.7
2	13.4	20.3	28.3	16.0	78.0	19.5
3	15.6	24.6	23.7	16.2	80.1	20.03
4	11.0	18.2	21.4	14.1	64.7	16.76
5	12.7	19.8	25.1	15.8	73.4	18.35
รวม ($T_{i.}$)	61.8	100.0	119.3	73.9	355.0	
ค่าเฉลี่ย	12.36	20.0	23.86	14.78		17.75

สมมติฐาน

H_o : ไม่มีความแตกต่างระหว่างยางทั้ง 4

H_o : $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.} = \mu$

H_o : $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

H_o' : ไม่มีความแตกต่างระหว่างเส้นทางทั้ง 5

H_o' : $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu_{.5} = \mu$

H_o' : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1 : มีความแตกต่างระหว่างยางทั้ง 4

H_1' : มีความแตกต่างระหว่างเส้นทางทั้ง 5

f_1 : $f\alpha,[(k-1),(k-1)(b-1)] = f_{0.01,(3,12)}$

$= 5.95$

f_2 : $f\alpha,[(b-1),(k-1)(b-1)] = f_{0.01,(4,12)}$

$= 5.41$

$$SSA = \sum_{i=1}^{k} \frac{T_{i.}^2}{b} - \frac{T_{..}^2}{N}$$

$$= \frac{61.8^2 + 100^2 + 119.3^2 + 73.9^2}{5} - \frac{355^2}{20}$$

$$= 401.338$$

$$SSB = \sum_{j=1}^{b} \frac{T_{.j}^2}{k} - \frac{T_{..}^2}{N}$$

$$= \frac{58.8^2 + 78^2 + 80.1^2 + 64.7^2 + 73.4^2}{4} - \frac{355^2}{20}$$

$$= 81.525$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij}^2 - \frac{T_{..}^2}{N}$$

$$= 9.1^2 + 17.1^2 + \ldots\ldots + 25.1^2 + 14.78^2 - \frac{355^2}{20}$$

$$= 6{,}810.28 - \frac{355^2}{25} = 509.03$$

SSE = SST - SSA - SSB

= 6,810.28 - 401.338 - 81.525

= 26.167

SOURCE	DF	SS	MS	f	f_α
(A) Treatment	3	401.338	133.779	61.340	5.95
(B) Block	4	81.525	20.381	9.345	5.41
(E) Error	12	26.167	2.181		
(T) Total	19	509.03			

Reject H_o และ H_o' ที่ $\alpha = 0.01$

7.7 แบบฝึกหัด

1. ในการศึกษากำลังรับแรงอัดของคอนกรีต บริษัทได้เปรียบเทียบกระบวนการผลิตที่แตกต่างกัน 4 วิธี โดยทำการเก็บรวบรวมข้อมูลไว้ดังต่อไปนี้

กระบวนการผลิต	กำลังรับแรงอัดของคอนกรีต (ksc)			
1	220	211	201	204
2	225	233	210	222
3	197	204	210	215
4	183	190	183	195

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่ากระบวนการผลิตที่แตกต่างกันมีผลกระทบต่อกำลังรับแรงอัดของคอนกรีตหรือไม่

2. ในการศึกษาว่าอุณหภูมิความร้อนที่ใช้ในการเผาอิฐมีผลต่อความหนาแน่นของอิฐชนิดหนึ่งหรือไม่ ได้ทำการทดลองโดยใช้อุณหภูมิในการเผาที่แตกต่างกัน 4 ระดับ เก็บข้อมูลได้ดังต่อไปนี้

อุณหภูมิ ($^{\circ}$C)	ความหนาแน่น				
100	21.8	21.9	21.7	21.6	21.7
125	21.7	21.4	21.5	21.4	
150	21.9	21.8	21.6	21.5	21.8
175	21.9	21.7	21.8	21.4	

จงทดสอบว่าอุณหภูมิที่ใช้ในการเผามีผลต่อความหนาแน่นของอิฐชนิดนี้หรือไม่ ระดับนัยสำคัญ 0.05

3. ในการเปรียบเทียบการออกแบบผังวงจรคอมพิวเตอร์ที่แตกต่างกัน 4 แบบ ว่ามีความแตกต่างกันหรือไม่ บริษัทได้ทำการทดลองเก็บรวบรวมข้อมูลปริมาณสิ่งรบกวน (noise) ดังตาราง

แบบวงจร	ปริมาณสิ่งรบกวน				
1	19	20	19	30	8
2	80	61	73	56	80
3	47	26	25	35	50
4	95	46	83	78	97

จากข้อมูลที่ได้นี้ จะสรุปได้หรือไม่ว่าค่าเฉลี่ยปริมาณสิ่งรบกวนของแบบผังวงจรทั้ง 4 มีความแตกต่างที่ระดับนัยสำคัญ 0.01

4. ในการทดลองเพื่อเปรียบเทียบวงจรไฟฟ้า 3 แบบ ที่จะนำมาใช้ในระบบเปิดปิดวาล์วอัตโนมัติได้ทำการวัดเวลาตอบสนอง (response time, milliseconds) ผลการทดลองเป็นดังนี้

ชนิดของวงจร	เวลาตอบสนอง				
A	9	12	10	8	15
B	20	21	23	17	30
C	6	5	8	16	7

ที่ระดับนัยสำคัญ 0.01 จงทดสอบสมมติฐานที่ว่า วงจรไฟฟ้าทั้ง 3 แบบมีเวลาตอบสนองเท่ากัน

5. ในการเปรียบเทียบอายุการใช้งาน (สัปดาห์) ของแบตเตอรี่ 3 ยี่ห้อ ได้ทำการเก็บรวบรวมข้อมูลอายุการใช้งานของแบตเตอรี่ทั้งสามยี่ห้อนั้น โดยทำการสุ่มตัวอย่างแบตเตอรี่ยี่ห้อละ 5 ตัวอย่าง บันทึกผลดังตาราง

อายุการใช้งานของแบตเตอรี่ (สัปดาห์)

ยี่ห้อที่ 1	ยี่ห้อที่ 2	ยี่ห้อที่ 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

จงวิเคราะห์ความแปรปรวนที่ระดับนัยสำคัญ 0.01 เพื่อทดสอบสมมติฐานว่า อายุการใช้งานเฉลี่ยของแบตเตอรี่ทั้งสามยี่ห้อไม่มีความแตกต่าง

6. ในการทดลองเพื่อศึกษาดูผลกระทบของอัตราการป้อนชิ้นงานในการใช้เครื่อง CNC เพื่อผลิตชิ้นส่วนซึ่งใช้ในเครื่องบินชนิดหนึ่ง วิศวกรผู้ผลิตซึ่งรับผิดชอบในการทดสอบ ต้องการทราบว่าขนาดของชิ้นส่วนที่ผลิตซึ่งเป็นชิ้นส่วนที่สำคัญยิ่งในเครื่องบินชนิดนี้นั้นขึ้นอยู่กับอัตราการป้อนชิ้นงานให้แก่เครื่อง CNC หรือไม่ ซึ่งจากข้อมูลในอดีตพบว่าการเปลี่ยนอัตราการป้อนชิ้นงานไม่มีผลกระทบต่อขนาดเฉลี่ยของชิ้นงานแต่อย่างใด หากแต่อาจมีผลกระทบต่อการเปลี่ยนแปลง (Variability) ของรูปแบบ วิศวกรจึงทำการทดลองใหม่โดยใช้อัตราการป้อนชิ้นงานที่แตกต่างกัน 4 ระดับ แต่ละระดับทำการผลิต 5 ครั้ง วัดส่วนเบี่ยงเบนมาตรฐานของขนาดชิ้นส่วน (10^{-3} มม.) ได้ข้อมูลดังตาราง

อัตราการป้อนชิ้นงาน (นิ้ว/นาที)	ส่วนเบี่ยงเบนมาตรฐานของขนาดชิ้นส่วน (x 10^{-3} มม.)				
10	0.09	0.10	0.13	0.08	0.07
12	0.06	0.09	0.12	0.07	0.12
14	0.11	0.08	0.08	0.05	0.06
15	0.19	0.13	0.15	0.25	0.11

จงทดสอบว่าอัตราการป้อนชิ้นงานมีผลกระทบต่อส่วนเบี่ยงเบนมาตรฐานของขนาดชิ้นส่วนที่สำคัญนี้หรือไม่ที่ระดับนัยสำคัญ 0.05

7. ครูสามคนได้ทำการสอนคณิตศาสตร์เบื้องต้นแก่นักเรียน 3 กลุ่ม ผลการสอบปลายภาคการศึกษาบันทึกไว้ดังต่อไปนี้

(หน่วย : คะแนน)

ครู		
ก	ข	ค
73	88	79
89	78	86
82	51	71
43	91	71
80	85	41
73	77	87
66	62	68
	76	59
	96	
	80	

จงทดสอบว่าคะแนนเฉลี่ยโดยครูทั้งสามคนมีความแตกต่างอย่างมีนัยสำคัญหรือไม่ที่ระดับนัยสำคัญ 0.05

8. ในการทดสอบคุณภาพของยางรถยนต์ 4 ชนิด ผู้ทดสอบได้นำยางรถทั้ง 4 ชนิดมาทดสอบกับรถ 3 คัน แล้วบันทึกระยะการใช้งานของยางรถยนต์ (x 10,000 กม.) ปรากฏผลดังนี้

รถคันที่ 1	รถคันที่ 2	รถคันที่ 3
ยางชนิดที่ 1 = 10.3	ยางชนิดที่ 2 = 7.5	ยางชนิดที่ 4 = 10.7
ยางชนิดที่ 2 = 10.5	ยางชนิดที่ 3 = 5.1	ยางชนิดที่ 1 = 9.6
ยางชนิดที่ 3 = 11.4	ยางชนิดที่ 1 = 7.7	ยางชนิดที่ 2 = 11.0
ยางชนิดที่ 4 = 11.2	ยางชนิดที่ 4 = 6.5	ยางชนิดที่ 3 = 11.5

จงทดสอบที่ระดับนัยสำคัญ 0.05 ว่าคุณภาพของยางแต่ละชนิดต่างกันหรือไม่ และรถทั้งสามคันนี้ใช้ยางเปลืองกว่ากันหรือไม่

9. ในการผลิตกระดาษของโรงงานแห่งหนึ่งได้ทำการทดสอบว่าสูตรผสมเยื่อกระดาษที่แตกต่างกัน 4 สูตร นั้น มีผลต่อคุณภาพของกระดาษหรือไม่ โดยที่โรงงานนี้เชื่อว่านอกจากสูตรผสมที่ต่างกันแล้ว พนักงานผู้ควบคุมการผลิตอาจมีผลกระทบต่อคุณภาพของกระดาษด้วย ข้อมูลในตารางคือ ความสามารถรับแรงดึงของกระดาษที่ผลิต

สูตรผสม \ พนักงาน	ก	ข	ค
A	114	120	117
B	126	119	123
C	137	127	134
D	141	129	127

จะสรุปได้หรือไม่ว่า สูตรผสมที่แตกต่างกันนั้นไม่มีผลต่อคุณภาพของกระดาษ และพนักงานผู้ควบคุมการผลิตมีประสิทธิภาพในการทำงานไม่แตกต่างกันที่ระดับนัยสำคัญ 0.01

10. ในการศึกษาเพื่อเปรียบเทียบวิตามิน 3 ชนิด ได้นำแฝดสามคนจาก 6 ครอบครัวมาเริ่มให้วิตามินตั้งแต่อายุ 1 ขวบ เด็กแต่ละคนถูกให้วิตามินชนิดหนึ่งอย่างสุ่มเป็นเวลา 2 ปี สมมติว่าข้อมูลข้างล่างนี้เป็นน้ำหนักที่เพิ่มขึ้น มีหน่วยเป็นปอนด์

ครอบครัว	วิตามิน			T_j	T_j^2
	A	B	C		
1	11.2	9.3	10.4	30.9	954.81
2	9.7	12.0	11.5	33.2	1,104.24
3	8.2	9.4	8.9	26.5	702.25
4	9.1	10.1	7.9	27.1	734.41
5	11.0	10.3	10.8	32.1	1,030.41
6	8.2	8.3	10.1	26.6	707.56
7	7.3	9.1	8.4	24.8	615.04
Ti	64.7	68.5	68.0		
T_i^2	4,186.9	4,692.25	4,624.00		
$\sum_i y_{ij}^2$	611.11	678.65	671.20		

$$T_{..} = 201.2, \quad \sum_i \sum_j y_{ij}^2 = 1{,}960.96$$

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่าวิตามินทั้งสามชนิดทำให้น้ำหนักที่เพิ่มขึ้นโดยเฉลี่ยมีความแตกต่างอย่างมีนัยสำคัญหรือไม่

บทที่ 8
การใช้คอมพิวเตอร์ช่วยในการวิเคราะห์ทางสถิติ
(Computer Aided Statistical Analysis)

8.1 คำนำ

Spreadsheet เช่น Excel ช่วยให้การวิเคราะห์ทางสถิติทำได้ง่ายขึ้น ในบทนี้จะได้กล่าวถึงแนวทางการใช้ Excel ช่วยในการวิเคราะห์ทางสถิติในเบื้องต้น อาทิเช่น Regression Analysis และ Analysis of Variances เพื่อให้ผู้อ่านเกิดความเข้าใจ และสามารถนำไปประยุกต์ใช้ต่อไป

8.2 ฟังก์ชันทางสถิติใน Excel

Excel มีฟังค์ชั่นสำหรับการวิเคราะห์หาค่าสถิติและวิเคราะห์ความน่าจะเป็นมากมาย อาทิเช่น

AVEDEV	ค่าเฉลี่ยของส่วนเบี่ยงเบนสัมบูรณ์ของชุดข้อมูลจากค่าเฉลี่ยข้อมูล
AVERAGE	ค่าเฉลี่ยของอาร์กิวเมนต์ทั้งหมด
AVERAGEA	ค่าเฉลี่ยของอาร์กิวเมนต์รวมทั้งตัวเลข ข้อความและค่าตรรกะ
BETADIST	ค่าการแจกแจงความน่าจะเป็นแบบเบต้าสะสม (cumulative beta probability density function)
BETAINV	ค่าผกผันของการแจกแจงความน่าจะเป็นแบบเบต้าสะสม (cumulative beta probability density function)
BINOMDIST	ค่าความน่าจะเป็นของการแจกแจงแบบทวินามสำหรับแต่ละชุดของผลการทดลอง
CHIDIST	ค่าความน่าจะเป็นด้านเดียวของการแจกแจงแบบไคสแควร์
CHIINV	ค่าผกผันของความน่าจะเป็นด้านเดียวของการแจกแจงแบบไคสแควร์
CHITEST	ค่าการทดสอบความเป็นอิสระ
CONFIDENCE	ค่าช่วงความเชื่อมั่น (confident interval) ของค่าเฉลี่ยประชากร
CORREL	ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างชุดข้อมูล 2 ชุดข้อมูล
COUNT	นับจำนวนที่อยู่ในรายการของอาร์กิวเมนต์ว่ามีเท่าไร
COUNTA	นับค่าที่อยู่ในรายการของอาร์กิวเมนต์ว่ามีเท่าไร

COVAR	ค่าความแปรปรวนร่วมซึ่งเป็นค่าเฉลี่ยของผลคูณของส่วนเบี่ยงเบนชนิดคู่
CRITBINOM	ค่าที่น้อยที่สุดที่ทำให้การแจกแจงแบบทวินามสะสมมีค่าน้อยกว่าหรือเท่ากับค่าเกณฑ์
DEVSQ	ค่าผลรวมยกกำลังสองของส่วนเบี่ยงเบน
EXPONDIST	ค่าการแจกแจงแบบเอ็กซ์โพเนนเชียล
FDIST	ค่า F ของการแจกแจงความน่าจะเป็น
FINV	ค่าผกผันของค่า F ของการแจกแจงความน่าจะเป็น
FISHER	ค่าการแปลง Fisher
FISHERINV	ค่าผกผันของการแปลง Fisher
FORECAST	ค่าตามแนวโน้มเชิงเส้น
FREQUENCY	การแจกแจงความถี่เป็นอาร์เรย์แนวตั้ง
FTEST	ค่าผลลัพธ์ของการทดสอบ F (F-test)
GAMMADIST	ค่าการแจกแจงแบบแกมมา
GAMMAINV	ค่าผกผันของการแจกแจงแบบแกมมาสะสม (gamma cumulative distribution)
GAMMALN	ค่าลอการิทึมธรรมชาติของฟังก์ชันแกมมา ?(x)
GEOMEAN	ค่าเฉลี่ยเรขาคณิต
GROWTH	ค่าตามเส้นแนวโน้มเอ็กซ์โพเนนเชียล
HARMEAN	ค่าเฉลี่ยฮาร์มอนิก
HYPGEOMDIST	ค่าการแจกแจงแบบไฮเพอร์ยีออเมตริก
INTERCEPT	ส่วนตัดของเส้นถดถอยเชิงเส้น
KURT	ค่าเคอร์โทซิส (kurtosis) ของชุดข้อมูล
LARGE	ค่าที่มากที่สุดลำดับที่ k-th ในชุดข้อมูล
LINEST	พารามิเตอร์ของแนวโน้มเชิงเส้น
LOGEST	พารามิเตอร์ของเส้นแนวโน้มเอ็กซ์โพเนนเชียล
LOGINV	ค่าผกผันของการแจกแจงแบบ lognormal
LOGNORMDIST	ค่าการแจกแจงแบบ lognormal สะสม (cumulative lognormal distribution)
MAX	ค่าที่มากที่สุดในรายการของอาร์กิวเมนต์
MAXA	ค่าที่มากที่สุดในรายการของอาร์กิวเมนต์รวมทั้งตัวเลข ข้อความและค่าตรรกะ
MEDIAN	ค่ามัธยฐานของตัวเลขที่ระบุ

MIN	ค่าที่น้อยที่สุดในรายการของอาร์กิวเมนต์
MINA	ค่าที่น้อยที่สุดในรายการของอาร์กิวเมนต์ รวมทั้งตัวเลข ข้อความและค่าตรรกะ
MODE	ค่าฐานนิยมในชุดข้อมูล
NEGBINOMDIST	ค่าการแจกแจงแบบทวินามลบ (negative binomial distribution)
NORMDIST	ค่าการแจกแจงแบบปกติสะสม
NORMINV	ค่าผกผันของการแจกแจงแบบปกติสะสม
NORMSDIST	ค่าการแจกแจงแบบปกติมาตรฐานสะสม
NORMSINV	ค่าผกผันของการแจกแจงแบบปกติมาตรฐานสะสม
PEARSON	ค่าสัมประสิทธิ์สหสัมพันธ์ของผลิตภัณฑ์เพียร์สัน (Pearson product moment correlation)
PERCENTILE	ส่งกลับค่าเปอร์เซ็นต์ไทล์ที่ k ของค่าในช่วงที่ระบุ
PERCENTRANK	ส่งกลับค่าลำดับในรูปเปอร์เซ็นต์ของค่าในชุดข้อมูลที่ระบุ
PERMUT	จำนวนวิธีเรียงสับเปลี่ยนสำหรับจำนวนวัตถุที่กำหนด
POISSON	ค่าการแจกแจงแบบปัวซอง
PROB	ค่าความน่าจะเป็นที่ค่าในช่วงจะอยู่ระหว่างสองขีดจำกัด (limit)
QUARTILE	ค่าควอร์ไทล์ (quartile) ของชุดข้อมูล
RANK	ลำดับที่ของตัวเลขในรายการของตัวเลข
RSQ	ค่ากำลังสองของสัมประสิทธิ์สหสัมพันธ์ชั่วขณะของผลิตภัณฑ์เพียร์สัน (Pearson product moment correlation coefficient)
SKEW	ค่าความเบ้ของการแจกแจง
SLOPE	ค่าความชันของการถดถอยเชิงเส้น
SMALL	ค่าที่น้อยที่สุดของลำดับที่ k ในชุดข้อมูล
STANDARDIZE	ค่ามาตรฐาน (normalized value)
STDEV	วิเคราะห์หาส่วนเบี่ยงเบนมาตรฐานที่มีพื้นฐานอยู่บนค่าตัวอย่าง
STDEVA	ประมาณค่าส่วนเบี่ยงเบนมาตรฐานที่มีพื้นฐานอยู่บนค่าตัวอย่างโดยให้รวมตัวเลข ข้อความและค่าตรรกะ
STDEVP	คำนวณหาส่วนเบี่ยงเบนมาตรฐาน โดยขึ้นอยู่กับประชากรทั้งหมด
STDEVPA	คำนวณหาส่วนเบี่ยงเบนมาตรฐานจากโดยขึ้นอยู่กับประชากรทั้งหมด โดยให้รวมตัวเลข ข้อความและค่าตรรกะ

STEYX	ค่าความคลาดเคลื่อนมาตรฐานของค่า y ที่ถูกทำนายสำหรับค่า x แต่ละค่าบนเส้นถดถอย
TDIST	ค่า t ในรูปเปอร์เซ็นต์ของการแจกแจงค่า t (student)
TINV	ค่าผกผันของการแจกแจงที (student)
TREND	ค่าตามเส้นแนวโน้มเชิงเส้น
TRIMMEAN	ค่ามัชฌิม(ค่าเฉลี่ย) ของชุดข้อมูลที่เหลือ
TTEST	ค่าความน่าจะเป็นที่ได้จากการทำการทดสอบ t (students' t-test)
VAR	ประมาณค่าความแปรปรวนโดยใช้ค่าตัวอย่างเป็นพื้นฐาน
VARA	ประมาณหาค่าความแปรปรวนโดยใช้ค่าตัวอย่างเป็นพื้นฐานโดยให้รวมตัวเลขข้อความและค่าตรรกะ
VARP	คำนวณหาค่าความแปรปรวนโดยใช้ประชากรทั้งหมดในการคำนวณ
VARPA	คำนวณหาค่าความแปรปรวนโดยใช้ประชากรทั้งหมดในการคำนวณโดยให้รวมตัวเลขข้อความและค่าตรรกะ
WEIBULL	ค่าการแจกแจงแบบ Weibull
ZTEST	ค่า P สองด้านของการทดสอบ z

8.3 การใช้ Spreadsheet ในการวิเคราะห์การถดถอยและความแปรปรวน

Spreadsheet เช่น Excel มีเครื่องมือช่วยในการวิเคราะห์ทางสถิติที่มีความซับซ้อน เช่น การวิเคราะห์การถดถอย (Regression Analysis) และการวิเคราะห์ความแปรปรวน (Analysis of Variance) หรือ ANOVA โดยการเรียก Data Analysis จากเมนูเครื่องมือ(Tool) ถ้าไม่พบ Data Analysis ให้เข้าเมนูเครืองมือ แล้ว Add-Ins Analysis ToolPak หรือ Analysis ToolPak - VBA ก็ได้

ใน Data Analysis มีเครื่องมือสำหรับการวิเคราะห์ทางสถิติ ได้แก่

1. Anova: Single Factor
2. Anova: Two-Factor With Replication
3. Anova: Two-Factor Without Replication
4. Correlation
5. Covariance
6. Descriptive Statistics

7. Exponential Smoothing
8. F-Test Two-Sample for Variances
9. Fourier Analysis
10. Histogram
11. Moving Average
12. Random Number Generation
13. Rank and Percentile
14. Regression
15. Sampling
16. t-Test: Paired Two Sample For Means
17. t-Test: Two-Sample Assuming Equal Variances
18. t-Test: Two-Sample Assuming Unequal Variances
19. z-Test

ในหัวข้อนี้ จะยกตัวอย่างเฉพาะการใช้เครื่องมือช่วยในการวิเคราะห์สหสัมพันธ์ การวิเคราะห์การถดถอย และการวิเคราะห์ความแปรปรวน

ตัวอย่างที่ 8.1 จงวิเคราะห์สหสัมพันธ์ของตัวแปรและวิเคราะห์การถดถอย จากข้อมูลในตารางที่กำหนดให้

Y	X1	X2	X3	X4
10	6	1	3	4
18	10	5	6	7
29	15	4	8	10
37	20	7	10	11
45	25	9	12	15
51	30	12	15	20
60	35	14	18	23
80	40	15	21	30

ผลการวิเคราะห์สหสัมพันธ์ โดยใช้ Correlationได้ Correlation Matrix ดังตาราง

	Y	X1	X2	X3	X4
Y	1.0000				
X1	0.9890	1.0000			
X2	0.9592	0.9821	1.0000		
X3	0.9912	0.9963	0.9838	1.0000	
X4	0.9890	0.9843	0.9670	0.9919	1.0000

ผลการวิเคราะห์การถดถอยโดยใช้ Regressionได้ผลดังตาราง

Regression Statistics	
Multiple R	0.9959
R Square	0.9918
Adjusted R Square	0.9808
Standard Error	3.1667
Observations	8

ANOVA

	df	SS	MS	F	Significance F
Regression	4	3617.42	904.35	90.18	0.0019
Residual	3	30.08	10.03		
Total	7	3647.50			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-2.832	3.824	-0.741	0.513	-15.001	9.337
X1	0.824	1.241	0.664	0.554	-3.125	4.772
X2	-2.107	1.443	-1.461	0.240	-6.698	2.484
X3	2.987	3.867	0.773	0.496	-9.318	15.292
X4	0.558	1.237	0.451	0.683	-3.380	4.496

Regression Model :

Y = -2.832 + 0.824X1 - 2.107X2 + 2.987X3 + 0.558X4

ตัวอย่างที่ 8.2 จงวิเคราะห์ความแปรปรวนเพื่อทดสอบว่าคะแนนสอบของนิสิตชายและนิสิตหญิงมีความแตกต่างกันทางสถิติที่ระดับนัย 0.05 จากข้อมูลในตารางที่กำหนดให้

	คะแนนสอบ (%)	
	นิสิตชาย	นิสิตหญิง
1	60	65
2	64	55
3	70	77
4	90	85

ใช้ **Anova: Single Factor**

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Male	4	284	71	177.3333
Female	4	282	70.5	174.3333

ANOVA

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.5	1	0.5	0.002844	0.959204	5.987374
Within Groups	1055	6	175.8333			
Total	1055.5	7				

ตัวอย่างที่ 8.3 จงวิเคราะห์ความแปรปรวนเพื่อทดสอบว่าคะแนนสอบของนิสิตชายและนิสิตหญิงของภาควิชาวิศวกรรมเกษตร วิศวกรรมชลประทาน และวิศวกรรมการอาหาร มีความแตกต่างกันทางสถิติที่ระดับนัย 0.05 จากข้อมูลในตารางที่กำหนดให้

	คะแนนสอบ (%)	
	นิสิตชาย	นิสิตหญิง
AE	70	85
	75	89
	70	45
	20	91
IRRE	45	70
	50	30
	75	45
	39	55
FE	71	57
	90	56
	30	45
	55	31

ใช้ ANOVA - Two Factors with Replication

SUMMARY	Male	Female	Total
AE			
Count	4	4	8
Sum	235	310	545
Average	58.75	77.5	68.125
Variance	672.92	475.67	592.70
IRRE			
Count	4	4	8
Sum	209	200	409
Average	52.25	50	51.125
Variance	250.25	283.33	230.13
FE			
Count	4	4	8
Sum	246	189	435
Average	61.50	47.25	54.38
Variance	645.67	146.92	397.70

Total		
Count	12	12
Sum	690	699
Average	57.50	58.25
Variance	444.27	450.57

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample(Department)	1303	2	651.50	1.58	0.23	3.55
Columns(Sex)	3.375	1	3.38	0.01	0.93	4.41
Interaction(Sex.Dept)	1116	2	558.00	1.35	0.28	3.55
Within	7424.25	18	412.46			
Total	9846.63	23				

8.4 แบบฝึกหัด

จงหาข้อมูลผลการทดลอง หรือ ผลการสำรวจ แล้ววิเคราะห์ทางสถิติ ดังต่อไปนี้

(1) Simple Regression

(2) Multiple Regression + Correlation Matrix

(3) ANOVA - single factor

(4) ANOVA - two factors with replication

บรรณานุกรม

1. ศันสนีย์ สุภาภา. (2539). ความน่าจะเป็นและสถิติประยุกต์สำหรับวิศวกร. ฟิสิกซ์เซ็นเตอร์. กรุงเทพฯ. 390 น.

2. วราวุธ วุฒิวณิชย์. (2544). เอกสารคำสอนวิชา 206221 ความน่าจะเป็นและสถิติประยุกต์สำหรับวิศวกร. ภาควิชาวิศวกรรมชลประทาน คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน. 318 น.

2. Hogg, R.V. (1989). Engineering Statistics, Macmillan Publishing Company, USA.

3. Johnson, H.A. and G.K. Bhattacharya (1992). Statistics-Principles and Methods (2nd ed.), John Wiley & Sons Inc., USA, 686 p.

4. Walpole, R.E. and R.H. Myers (1993). Probability and Statistics for Engineering and Scientists (5th ed.). Macmillan Publishing Company, USA. 776 p.

ภาคผนวก

Table A1 Binomial Sum ,B(r;n,p)

N	r	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0000	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A2 Poisson Cumulative Probability , P(r;μ)

	μ									
r	0.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
0	0.6065	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001
1	0.9098	0.7358	0.4060	0.1991	0.0916	0.0404	0.0174	0.0073	0.0030	0.0012
2	0.9856	0.9197	0.6767	0.4232	0.2381	0.1247	0.0620	0.0296	0.0138	0.0062
3	0.9982	0.9810	0.8571	0.6472	0.4335	0.2650	0.1512	0.0818	0.0424	0.0212
4	0.9998	0.9963	0.9473	0.8153	0.6288	0.4405	0.2851	0.1730	0.0996	0.0550
5	1.0000	0.9994	0.9834	0.9161	0.7851	0.6160	0.4457	0.3007	0.1912	0.1157
6	1.0000	0.9999	0.9955	0.9665	0.8893	0.7622	0.6063	0.4497	0.3134	0.2068
7	1.0000	1.0000	0.9989	0.9881	0.9489	0.8666	0.7440	0.5987	0.4530	0.3239
8	1.0000	1.0000	0.9998	0.9962	0.9786	0.9319	0.8472	0.7291	0.5925	0.4557
9	1.0000	1.0000	1.0000	0.9989	0.9919	0.9682	0.9161	0.8305	0.7166	0.5874
10	1.0000	1.0000	1.0000	0.9997	0.9972	0.9863	0.9574	0.9015	0.8159	0.7060
11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9945	0.9799	0.9467	0.8881	0.8030
12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9980	0.9912	0.9730	0.9362	0.8758
13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9964	0.9872	0.9658	0.9261
14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9943	0.9827	0.9585
15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9976	0.9918	0.9780
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9963	0.9889
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9947
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9976
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9989
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996

Table A2(cont') Poisson Cumulative Probability , $P(r;\mu)$

	μ									
r	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0103	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000
4	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001	0.0000
5	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003	0.0002
6	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010	0.0005
7	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029	0.0015
8	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071	0.0039
9	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154	0.0089
10	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304	0.0183
11	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549	0.0347
12	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917	0.0606
13	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426	0.0984
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081	0.1497
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867	0.2148
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751	0.2920
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686	0.3784
18	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622	0.4695
19	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509	0.5606
20	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307	0.6472
21	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991	0.7255
22	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551	0.7931
23	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989	0.8490
24	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317	0.8933
25	1.0000	0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554	0.9269
26	1.0000	1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718	0.9514
27	1.0000	1.0000	0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827	0.9687
28	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897	0.9805
29	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9989	0.9973	0.9941	0.9882
30	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9967	0.9930

Table A3 Area under normal curve , P(Z≤ z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table A3 (cont') Area under normal curve , P(Z≤ z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table A4 Critical Values of the χ^2-Distribution

$\alpha = P[\chi2 > \chi2(\alpha,\nu)]$

	α									
ν	0.995	0.99	0.98	0.975	0.95	0.90	0.80	0.75	0.70	0.50
1	0.000	0.000	0.001	0.001	0.004	0.016	0.064	0.102	0.148	0.455
2	0.010	0.020	0.040	0.051	0.103	0.211	0.446	0.575	0.713	1.386
3	0.072	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	25.336
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	26.336
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	27.336
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	28.336
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	29.336

Table A4 (cont') Critical Values of the χ^2-Distribution

$\alpha = P[\chi2 > \chi2(\alpha,\nu)]$

ν	α 0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702

Table A5 Random Numbers

Line/Col.	1	2	3	4	5	6	7	8	9	10
1	06443	02305	55491	94052	06440	70956	84536	48977	23549	92412
2	97405	39281	99337	12649	45957	66462	46316	87497	27451	15694
3	54494	09114	25175	62811	91965	08809	39199	95634	44756	70018
4	71717	55136	15321	85617	61193	98908	98987	73718	08464	43363
5	92732	41544	20457	24963	86482	59783	50806	90184	20992	83129
6	34930	67309	58211	07045	11470	70101	82709	09190	73183	50949
7	96159	94595	34926	28319	37426	77907	46334	69994	31333	44004
8	85112	05712	15057	93493	12182	58752	20820	90741	88135	91576
9	85543	37187	32280	43432	65923	47323	79147	23155	27783	81558
10	99853	73924	29586	83331	10354	8232	67304	16165	79365	64123
11	71459	23891	54522	84708	42998	06334	22768	25852	10679	98927
12	83544	11617	09245	15339	68106	90531	23803	89935	73761	32008
13	70186	21938	77958	51333	58521	57122	13791	65442	24742	07550
14	72459	37509	99447	91526	99248	39775	49447	37749	57730	90374
15	84212	67351	72803	18425	75997	84690	63976	25175	66522	30288
16	77484	91600	19823	99429	12736	98055	15620	30009	35410	86417
17	23642	05814	24857	43167	66249	56640	81692	20306	44194	52259
18	53065	76052	23722	99779	48614	58615	78389	76609	93714	84540
19	89770	49112	48785	35948	61678	56361	47044	42036	56798	31919
20	98899	8500	98182	65453	35513	90417	06400	73412	96591	38465
21	84346	82206	62166	52467	01586	96048	05606	77506	11551	59992
22	05538	10158	83592	76858	45990	90984	71424	72947	33155	41296
23	89885	81060	13850	37338	18583	28483	54653	68890	88587	33537
24	86475	91656	78057	22394	48415	05282	94962	42855	95367	19964
25	92162	45059	47714	18203	93775	85054	35341	06641	47295	07702
26	63511	60927	89151	76801	83603	82111	34979	54016	13473	40123
27	65190	52491	35924	29651	22833	67571	33495	75446	81142	14834
28	10092	30796	67899	57256	35792	29940	66238	39723	40777	89479
29	92739	09784	12644	97170	28617	65313	63693	46393	44687	88750
30	34947	33811	76659	86525	86757	63590	34838	63460	57211	88842
31	35013	81975	94284	85001	15421	73699	23360	42649	75379	16372
32	06290	92162	40211	35257	83486	37570	95690	96182	29356	28314
33	88401	07201	46498	74264	69774	68223	41464	43327	30771	20258
34	71774	18063	22563	09617	69989	79224	75706	24911	69581	76894
35	95767	49158	01785	54159	16933	24906	26079	44764	52270	62196
36	47700	86202	07250	26131	56641	90478	04059	82819	88361	09959
37	30189	12118	36339	10661	27365	75838	25673	52049	92517	31386
38	63867	61042	19121	09432	54307	65398	90566	72745	50102	68741
39	26084	90519	09102	13350	31599	59658	19920	57399	39167	84749
40	44397	42586	59942	07598	42703	02882	26472	01305	33751	89612
41	05351	73952	00494	68703	52165	34813	21544	08683	97563	91993
42	80840	52783	06540	46097	36537	48958	37823	84037	52106	53995
43	04247	27870	30723	52655	45101	55884	61899	31265	71734	20392
44	92768	00136	60131	69212	48991	61666	40936	53844	24919	50447
45	56324	83704	77054	72769	88113	74640	60562	00064	48245	38978
46	26957	68024	89403	50456	53010	33961	55333	43402	72602	49096
47	92211	20953	02723	59741	38868	46760	48167	71675	39674	57074
48	49063	65712	74858	03102	87572	65166	12031	75964	62812	96093
49	50858	93905	60114	05431	63546	02486	36136	16639	31324	58215
50	54033	76851	75219	46747	22625	19185	37485	60095	12282	43604

Table A6 Critical Values of the t-Distribution

$\alpha = P[T>t(\alpha,\nu)]$

	α											
ν	0.4	0.3	0.2	0.1	0.05	0.025	0.02	0.015	0.01	0.005	0.0025	0.0005
1	0.325	0.727	1.376	3.078	6.314	12.706	15.894	21.205	31.821	63.656	127.321	636.578
2	0.289	0.617	1.061	1.886	2.920	4.303	4.849	5.643	6.965	9.925	14.089	31.600
3	0.277	0.584	0.978	1.638	2.353	3.182	3.482	3.896	4.541	5.841	7.453	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	2.999	3.298	3.747	4.604	5.598	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	2.757	3.003	3.365	4.032	4.773	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	2.612	2.829	3.143	3.707	4.317	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	2.517	2.715	2.998	3.499	4.029	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	2.449	2.634	2.896	3.355	3.833	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	2.398	2.574	2.821	3.250	3.690	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	2.359	2.527	2.764	3.169	3.581	4.587
11	0.260	0.540	0.876	1.363	1.796	2.201	2.328	2.491	2.718	3.106	3.497	4.437
12	0.259	0.539	0.873	1.356	1.782	2.179	2.303	2.461	2.681	3.055	3.428	4.318
13	0.259	0.538	0.870	1.350	1.771	2.160	2.282	2.436	2.650	3.012	3.372	4.221
14	0.258	0.537	0.868	1.345	1.761	2.145	2.264	2.415	2.624	2.977	3.326	4.140
15	0.258	0.536	0.866	1.341	1.753	2.131	2.249	2.397	2.602	2.947	3.286	4.073
16	0.258	0.535	0.865	1.337	1.746	2.120	2.235	2.382	2.583	2.921	3.252	4.015
17	0.257	0.534	0.863	1.333	1.740	2.110	2.224	2.368	2.567	2.898	3.222	3.965
18	0.257	0.534	0.862	1.330	1.734	2.101	2.214	2.356	2.552	2.878	3.197	3.922
19	0.257	0.533	0.861	1.328	1.729	2.093	2.205	2.346	2.539	2.861	3.174	3.883
20	0.257	0.533	0.860	1.325	1.725	2.086	2.197	2.336	2.528	2.845	3.153	3.850
21	0.257	0.532	0.859	1.323	1.721	2.080	2.189	2.328	2.518	2.831	3.135	3.819
22	0.256	0.532	0.858	1.321	1.717	2.074	2.183	2.320	2.508	2.819	3.119	3.792
23	0.256	0.532	0.858	1.319	1.714	2.069	2.177	2.313	2.500	2.807	3.104	3.768
24	0.256	0.531	0.857	1.318	1.711	2.064	2.172	2.307	2.492	2.797	3.091	3.745
25	0.256	0.531	0.856	1.316	1.708	2.060	2.167	2.301	2.485	2.787	3.078	3.725
26	0.256	0.531	0.856	1.315	1.706	2.056	2.162	2.296	2.479	2.779	3.067	3.707
27	0.256	0.531	0.855	1.314	1.703	2.052	2.158	2.291	2.473	2.771	3.057	3.689
28	0.256	0.530	0.855	1.313	1.701	2.048	2.154	2.286	2.467	2.763	3.047	3.674
29	0.256	0.530	0.854	1.311	1.699	2.045	2.150	2.282	2.462	2.756	3.038	3.660
30	0.256	0.530	0.854	1.310	1.697	2.042	2.147	2.278	2.457	2.750	3.030	3.646

Table A7-1 Critical Values of the f-Distribution

α= P[F>f(α, ν1 ,ν2)] =0.05

ν2	ν1 = 1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
100000	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Table A7-1(cont') Critical Values of the f-Distribution

$\alpha = P[F>f(\alpha, \nu 1, \nu 2)] = 0.05$

	ν1									
ν2	10	12	15	20	24	30	40	60	120	100000
1	241.88	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
100000	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.01

Table A7-2 Critical Values of the f-Distribution

α= P[F>f(α, ν1 ,ν2)] =0.01

	ν1								
ν2	1	2	3	4	5	6	7	8	9
1	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95	6022.40
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
100000	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Table A7-2(cont') Critical Values of the f-Distribution

α= P[F>f(α, ν1 ,ν2)] =0.01

	ν1									
ν2	10	12	15	20	24	30	40	60	120	100000
1	6055.93	6106.68	6156.97	6208.66	6234.27	6260.35	6286.43	6312.97	6339.51	6365.59
2	99.40	99.42	99.43	99.45	99.46	99.47	99.48	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
100000	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.01

Table A8 Sample size(n) for testing mean by t test

	Level of t-test																				
Single-sided	α=0.005					α=0.01					α=0.025					α=0.05					
Double-sided	α=0.01					α=0.02					α=0.05					α=0.1					
β=	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
0.05																					0.05
0.10																					0.10
0.15																				122	0.15
0.20										139					99					70	0.20
0.25					110					90				128	64			139	101	45	0.25
0.30				134	78				115	63			119	90	45		122	97	71	32	0.30
0.35			125	99	58			109	85	47		109	88	67	34		90	72	52	24	0.35
0.40		115	97	77	45		101	85	66	37	117	84	68	51	26	101	70	55	40	19	0.40
0.45		92	77	62	37	110	81	68	53	30	93	67	54	41	21	80	55	44	33	15	0.45
0.50	100	75	63	51	30	90	66	55	43	25	76	54	44	34	18	65	45	36	27	13	0.50
Value of Δ=(μ−μo)/σ																					
0.55	83	63	53	42	26	75	55	46	36	21	63	45	37	28	15	54	38	30	22	11	0.55
0.60	71	53	45	36	22	63	47	39	31	18	53	38	32	24	13	46	32	26	19	9	0.60
0.65	61	46	39	31	20	55	41	34	27	16	46	33	27	21	12	39	28	22	17	8	0.65
0.70	53	40	34	28	17	47	35	30	24	14	40	29	24	19	10	34	24	19	15	8	0.70
0.75	47	36	30	25	16	42	31	27	21	13	35	26	21	16	9	30	21	17	13	7	0.75
0.80	41	32	27	22	14	37	28	24	19	12	31	22	19	15	9	27	19	15	12	6	0.80
0.85	37	29	24	20	13	33	25	21	17	11	28	21	17	13	8	24	17	14	11	6	0.85
0.90	34	26	22	18	12	29	23	19	16	10	25	19	16	12	7	21	15	13	10	5	0.90
0.95	31	24	20	17	11	27	21	18	14	9	23	17	14	11	7	19	14	11	9	5	0.95
1.00	28	22	19	16	10	25	19	16	13	9	21	16	13	10	6	18	13	11	8	5	1.00
1.10	24	19	16	14	9	21	16	14	12	8	18	13	11	9	6	15	11	9	7		1.10
1.20	31	16	14	12	8	18	14	2	10	7	15	12	10	8	5	13	10	8	6		1.20
1.30	18	15	13	11	8	16	13	11	9	6	14	10	9	7		11	8	7	6		1.30
1.40	16	13	12	10	7	14	11	10	9	6	12	9	8	7		10	8	7	5		1.40
1.50	15	12	11	9	7	13	10	9	8	6	11	8	7	6		9	7	6			1.50
1.60	13	11	10	8	6	12	10	9	7	5	10	8	7	6		8	6	6			1.60
1.70	12	10	9	8	6	11	9	8	7		9	7	6	5		8	6	5			1.70
1.80	12	10	9	8	6	10	8	7	7		8	7	6			7	6				1.80
1.90	11	9	8	7	6	10	8	7	6		8	7	6			7	5				1.90
2.00	10	8	8	7	5	9	7	7	6		7	6	5			6					2.00
2.10	10	8	7	7		8	7	6	6		7	6				6					2.10
2.20	9	8	7	6		8	7	6	5		7	6				6					2.20
2.30	9	7	7	6		8	6	6			6	5				5					2.30
2.40	8	7	7	6		7	6	6			6										2.40
2.50	8	7	6	6		7	6	6			6										2.50
3.00	7	6	6	5		6	5	5			5										3.00
3.50	6	5	5				5														3.50
4.00	6																				4.00

Table A9 Barlett Statistic

b_k (0.01 ; n)

n	Number of Populations, k								
	2	3	4	5	6	7	8	9	10
3	0.1411	0.1672	*	*	*	*	*	*	*
4	0.2843	0.3165	0.3475	0.3729	0.3937	0.4110	*	*	*
5	0.3984	0.4304	0.4607	0.4850	0.5046	0.5207	0.5343	0.5458	0.5558
6	0.4850	0.5149	0.5430	0.5653	0.5832	0.5978	0.6100	0.6204	0.6293
7	0.5512	0.5787	0.6045	0.6248	0.6410	0.6542	0.6652	0.6744	0.6824
8	0.6031	0.6282	0.6518	0.6704	0.6851	0.6970	0.7069	0.7153	0.7225
9	0.6445	0.6676	0.6892	0.7062	0.7197	0.7305	0.7395	0.7471	0.7536
10	0.6783	0.6996	0.7195	0.7352	0.7475	0.7575	0.7657	0.7726	0.7786
11	0.7063	0.7260	0.7445	0.7590	0.7703	0.7795	0.7871	0.7935	0.7990
12	0.7299	0.7483	0.7654	0.7789	0.7894	0.7980	0.8050	0.8109	0.8160
13	0.7501	0.7672	0.7832	0.7958	0.8056	0.8135	0.8201	0.8256	0.8303
14	0.7674	0.7835	0.7985	0.8103	0.8195	0.8269	0.8330	0.8382	0.8426
15	0.7825	0.7977	0.8118	0.8229	0.8315	0.8385	0.8443	0.8491	0.8532
16	0.7958	0.8101	0.8235	0.8339	0.8421	0.8486	0.8541	0.8586	0.8625
17	0.8076	0.8211	0.8338	0.8436	0.8514	0.8576	0.8627	0.8670	0.8707
18	0.8181	0.8309	0.8429	0.8523	0.8596	0.8655	0.8704	0.8745	0.8780
19	0.8275	0.8397	0.8512	0.8601	0.8670	0.8727	0.8773	0.8811	0.8845
20	0.8360	0.8476	0.8586	0.8671	0.8737	0.8791	0.8835	0.8871	0.8903
21	0.8437	0.8548	0.8653	0.8734	0.8797	0.8848	0.8890	0.8926	0.8956
22	0.8507	0.8614	0.8714	0.8791	0.8852	0.8901	0.8941	0.8975	0.9004
23	0.8571	0.8673	0.8769	0.8844	0.8902	0.8949	0.8988	0.9020	0.9047
24	0.8630	0.8728	0.8820	0.8892	0.8948	0.8993	0.9030	0.9061	0.9087
25	0.8684	0.8779	0.8867	0.8936	0.8990	0.9034	0.9069	0.9099	0.9124
26	0.8734	0.8825	0.8911	0.8977	0.9029	0.9071	0.9105	0.9134	0.9158
27	0.8781	0.8869	0.8951	0.9015	0.9065	0.9105	0.9138	0.9166	0.9190
28	0.8824	0.8909	0.8988	0.9050	0.9099	0.9138	0.9169	0.9196	0.9219
29	0.8864	0.8946	0.9023	0.9083	0.9130	0.9167	0.9198	0.9224	0.9246
30	0.8902	0.8981	0.9056	0.9114	0.9159	0.9195	0.9225	0.9250	0.9271
40	0.9175	0.9235	0.9291	0.9335	0.9370	0.9397	0.9420	0.9439	0.9455
50	0.9339	0.9387	0.9433	0.9468	0.9796	0.9518	0.9536	0.9551	0.9564
60	0.9449	0.9489	0.9527	0.9557	0.9580	0.9599	0.9614	0.9626	0.9637
80	0.9586	0.9617	0.9646	0.9668	0.9685	0.9699	0.9711	0.9720	0.9728
100	0.9669	0.9693	0.9716	0.9734	0.9748	0.9759	0.9769	0.9776	0.9783

Table A9(cont') Barlett Statistic

b_k (0.05 ; n)

n	Number of Populations, k								
	2	3	4	5	6	7	8	9	10
3	0.3123	0.3058	0.3173	0.3299	*	*	*	*	*
4	0.4780	0.4699	0.4803	0.4921	0.5028	0.5122	0.5204	0.5277	0.5341
5	0.5845	0.5762	0.5850	0.5952	0.6045	0.6126	0.6197	0.6260	0.6315
6	0.6563	0.6483	0.6559	0.6646	0.6727	0.6798	0.6860	0.6914	0.6961
7	0.7075	0.7000	0.7065	0.7142	0.7213	0.7275	0.7329	0.7376	0.7418
8	0.7456	0.7387	0.7444	0.7512	0.7574	0.7629	0.7677	0.7719	0.7757
9	0.7751	0.7686	0.7737	0.7798	0.7854	0.7903	0.7946	0.7984	0.8017
10	0.7984	0.7924	0.7970	0.8025	0.8076	0.8121	0.8160	0.8194	0.8224
11	0.8175	0.8118	0.8160	0.8210	0.8257	0.8298	0.8333	0.8365	0.8392
12	0.8332	0.8280	0.8317	0.8364	0.8407	0.8444	0.8477	0.8506	0.8531
13	0.8465	0.8415	0.8450	0.8493	0.8533	0.8568	0.8598	0.8625	0.8648
14	0.8578	0.8532	0.8564	0.8604	0.8641	0.8673	0.8701	0.8726	0.8748
15	0.8676	0.8632	0.8662	0.8699	0.8734	0.8764	0.8790	0.8814	0.8834
16	0.8761	0.8719	0.8747	0.8782	0.8815	0.8843	0.8868	0.8890	0.8909
17	0.8836	0.8796	0.8823	0.8856	0.8886	0.8913	0.8936	0.8957	0.8975
18	0.8902	0.8865	0.8890	0.8921	0.8949	0.8975	0.8997	0.9016	0.9033
19	0.8961	0.8926	0.8949	0.8979	0.9006	0.9030	0.9051	0.9069	0.9086
20	0.9015	0.8980	0.9003	0.9031	0.9057	0.9080	0.9100	0.9117	0.9132
21	0.9063	0.9030	0.9051	0.9078	0.9103	0.9124	0.9143	0.9160	0.9175
22	0.9106	0.9075	0.9095	0.9120	0.9144	0.9165	0.9183	0.9199	0.9213
23	0.9146	0.9116	0.9135	0.9159	0.9182	0.9202	0.9219	0.9235	0.9248
24	0.9182	0.9153	0.9172	0.9195	0.9217	0.9236	0.9253	0.9267	0.9280
25	0.9216	0.9187	0.9205	0.9228	0.9249	0.9267	0.9283	0.9297	0.9309
26	0.9246	0.9219	0.9236	0.9258	0.9278	0.9296	0.9311	0.9325	0.9336
27	0.9275	0.9249	0.9265	0.9286	0.9305	0.9322	0.9337	0.9350	0.9361
28	0.9301	0.9276	0.9292	0.9312	0.9330	0.9347	0.9361	0.9374	0.9385
29	0.9326	0.9301	0.9316	0.9366	0.9354	0.9370	0.9383	0.9396	0.9406
30	0.9348	0.9325	0.9340	0.9358	0.9376	0.9391	0.9404	0.9416	0.9426
40	0.9513	0.9495	0.9506	0.9520	0.9533	0.9545	0.9555	0.9564	0.9572
50	0.9612	0.9597	0.9606	0.9617	0.9628	0.9637	0.9645	0.9652	0.9658
60	0.9677	0.9665	0.9672	0.9681	0.9690	0.9698	0.9705	0.9710	0.9716
80	0.9758	0.9749	0.9754	0.9761	0.9768	0.9774	0.9779	0.9783	0.9787
100	0.9807	0.9799	0.9804	0.9809	0.9815	0.9819	0.9823	0.9827	0.9830

Table A10 Cochran Statistic

α=.01

k \ n	2	3	4	5	6	7	8	9	10	11	17	37	145	∞
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8375	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.7270	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1250
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1881	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1566	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1608	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.3412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0867	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.0330	0.0957	0.0898	0.8530	0.0816	0.0668	0.5030	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.1225	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	0.0178	0.0125	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

α=.05

k \ n	2	3	4	5	6	7	8	9	10	11	17	37	145	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9056	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2022	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1505	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0