บทที่ 3

ทฤษฎีและสมมุติฐานที่นำมาใช้ในการวิจัย

3.1 นิยามเกี่ยวกับสถิติ (Statistical Definition)

ในการวิจัยครั้งนี้จะต้องมีความรู้ทางด้านสถิติเพื่อนำมาใช้ในการวิเคราะห์ความถี่ของข้อมูล

สถิติก็คือตัวเลข แต่เป็นตัวเลขจำนวนหนึ่งที่มีการรวบรวมขึ้น เพื่อให้ข้อเท็จจริงเกี่ยวกับลักษณะหรือคุณสมบัติของบางสิ่งบางอย่างที่แสดงออกเป็นตัวเลขได้ และจะต้องประกอบด้วยตัวเลขตั้งแต่สองตัวขึ้นไป และสถิติยังหมายถึงปริมาณโดยเฉพาะอย่าง เช่น ส่วนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน เลขดัชนี สัมประสิทธิ์ของสหสัมพันธ์ หรือจะกล่าวได้อีกแบบหนึ่งว่า สถิติก็คือ "ค่าที่แสดงออกถึงลักษณะของตัวอย่าง" แต่ "พารามิเตอร์" (Parameter) หมายถึง "ค่าที่แสดงออกถึงลักษณะของประชากร" หรือปริมาณโดยเฉพาะอย่างเช่น ส่วนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ซึ่งคำนวณได้จากข้อมูลทั้งหมดหรือประชากร

3.1.1 ประชากรและตัวอย่าง

วัตถุประสงค์ทางสถิติก็เพื่อที่จะบรรยายข้อเท็จจริงของกลุ่มข้อมูลให้บรรลุผลเท่าที่เราได้รายละเอียด จำนวนข้อมูลทั้งหมดหรือจำนวนข้อมูลที่อยู่ในช่วงการพิจารณาทั้งหมด เรียกว่า "ประชากร" "ข้อมูล" ไม่ได้หมายถึงเฉพาะคน สัตว์และสิ่งของเท่านั้น ยังหมายถึงจำนวนค่าสังเกต (Observation) ทั้งหมดก็ได้

ตัวอย่างหมายถึงส่วนหนึ่งของประชากรโดยไม่จำกัดเฉพาะว่าจะต้องเป็นส่วนใดส่วนหนึ่งหรือหมายถึงค่าสังเกตที่ประกอบขึ้น ประชากรจะเป็นจำนวนเท่าไรก็ได้ หรือในกรณีที่เลือกประชากรมาบางส่วน และใช้ส่วนของประชากรที่เลือกมานี้เป็นตัวแทนของประชากรทั้งหมด ส่วนของประชากรที่เลือกมานี้ เรียกว่า "ตัวอย่าง"

3.1.2 การวัดแนวโน้มเข้าสู่ส่วนกลาง

การแจกแจงความถี่ของข้อมูล มักมีส่วนคล้ายคลึงกัน 2 ประการคือ

1. ค่าของของข้อมูลในกลุ่มมีค่าแตกต่างกันออกไป (Variation)
2. แม้ว่ามีค่าแตกต่างกันออกไป ค่าเหล่านี้ยังมีแนวโน้มเอียงเข้าสู่ส่วนกลาง

จะเห็นได้จากข้อมูลส่วนใหญ่ตามข้อเท็จจริง เมื่อมีการแจกแจงโดยทั่ว ๆ ไป มักจับกลุ่มรวมกันอยู่ในแถบค่ากลาง ๆ ฉะนั้นในการเลือกข้อมูล เพื่อใช้เป็นตัวแทนของข้อมูลทั้งกลุ่ม ก็เป็นธรรมดาที่ควรจะเลือกใช้ค่ากลางหรือค่าที่เกิดขึ้นบ่อยที่สุด หรือค่าเฉลี่ย เพราะเป็นค่าที่ถือได้ว่า ใช้แทนข้อมูลทั้งกลุ่มได้ดีที่สุด

3.1.2.1 ส่วนเฉลี่ย (Mean)

ถ้า N เป็นขนาดของประชากร ประกอบด้วย $x_1, x_2 \ldots\ldots x_N$ ค่าเฉลี่ยของประชากรจะคำนวณได้จาก

$$\mu = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

ถ้า n เป็นขนาดของกลุ่มตัวอย่าง ประกอบด้วย $x_1, x_2 \ldots\ldots\ldots\ldots x_n$ ค่าเฉลี่ยของตัวอย่างจะคำนวณได้จากสูตร

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

3.1.2.2 ฐานนิยม (Mode)

หมายถึงค่าของข้อมูลที่เกิดบ่อยที่สุด ความถี่มากที่สุด

3.1.2.3 มัธยฐาน (Median)

หมายถึงข้อมูลตัวกลาง เมื่อเอาข้อมูลที่มีอยู่มาเรียงลำดับตามขนาดมากน้อย ถ้าจำนวนข้อมูลที่มีอยู่เป็นเลขคี่ ข้อมูลตัวกลางก็คือค่าของมัธยฐาน ถ้าจำนวนข้อมูลที่มีอยู่เป็นเลขคู่ ค่าเฉลี่ยของข้อมูลตัวกลางสองจำนวนก็คือค่าของมัธยฐาน

3.1.3 การกระจาย (Variation or Dispersion)

ส่วนเฉลี่ยมัธยฐาน เป็นลักษณะตัวแทนของข้อมูลกลุ่มหนึ่งเพียงลักษณะเดียว ส่วนลักษณะอื่น ๆ ของข้อมูล เช่น ความแตกต่างไม่อาจแสดงได้ด้วยส่วนเฉลี่ย ลักษณะของข้อมูลที่ส่วนเฉลี่ยไม่อาจแสดงได้นี้ มีความสำคัญไม่น้อยกว่าความโน้มเอียงเข้าสู่ส่วนกลาง จากสาเหตุนี้ในการเปรียบเทียบข้อมูลกลุ่มหนึ่งกับอีกกลุ่มหนึ่งจำเป็นต้องพิจารณาถึงการกระจายควบคู่ไปกับค่าเฉลี่ยของข้อมูลเสมอไป การวัดการกระจายอาจวัดเป็นหน่วยเดียวกันกับค่าของข้อมูล จัดว่า

เป็นการกระจายสมบูรณ์ ส่วนการวัดเป็นตัวเลขที่ปราศจากหน่วย เช่น เปอร์เซ็นต์จัดว่าเป็นการกระจายสัมพัทธ์ ในที่นี้จะขอกล่าว แต่เพียงการวัดการกระจายสมบูรณ์ (Absolute variability)

3.1.3.1 พิสัย (Range)

คือ ค่าผลต่างระหว่างค่าสูงสุดกับค่าต่ำสุดของข้อมูลในกลุ่ม

3.1.3.2 ส่วนเบี่ยงเบนเฉลี่ย (Average deviation)

คือ ส่วนเฉลี่ยแบบเลขคณิตของระยะข้อมูลทุกตัวในกลุ่มเบี่ยงเบนไปจากค่าส่วนเฉลี่ยของข้อมูลกลุ่มนั้น โดยไม่คำนึงถึงเครื่องหมายถ้ามีข้อมูลอยู่กลุ่มหนึ่ง $x_1, x_2, x_3 \ldots x_n$ ซึ่งมีส่วนเฉลี่ยเลขคณิต $\bar{X}$ ส่วนเบี่ยงเบนเฉลี่ยก็จะคำนวณได้จากสูตร

$$A.D. = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

3.1.3.3 ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

เป็นการวัดส่วนเบี่ยงเบนของข้อมูลแต่ละตัวอย่างจากส่วนเฉลี่ย ซึ่งคำนวณได้จากสูตร ดังนี้

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.4)$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.5)$$

เมื่อ σ เป็นค่าส่วนเบี่ยงเบนมาตรฐานของประชากร และ

s เป็นค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

3.1.3.4 ความแปรปรวน (Variance)

คือ ค่าส่วนเบี่ยงเบนจากกำลังสองเขียนได้เป็นสูตร ดังนี้

$$v = \sigma^2 \text{ หรือ } v = s^2 \qquad \ldots\ldots\ldots\ldots (3.6)$$

เมื่อ v เป็นค่าของความแปรปรวนสำหรับประชากรและตัวอย่าง

3.1.3.5 สัมประสิทธิ์ของความแปรปรวน (Coefficient of Variation)

คือ ค่าอัตราส่วนของค่าส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเขียนเป็นสูตรได้ดังนี้

$$C_v = \frac{\sigma}{\mu} \text{ หรือ } \frac{S}{\bar{x}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3.7)$$

เมื่อ C_v เป็นค่าสัมประสิทธิ์ของความแปรปรวนและเป็นค่าที่ไม่มีหน่วย

3.2 นิยามเกี่ยวกับความน่าจะเป็น (Probability Definition)

บ่อยครั้งที่ความน่าจะเป็นมักจะถูกกล่าวด้วยความไม่มั่นใจ ทฤษฎีทางคณิตศาสตร์ของความน่าจะเป็น สำหรับแซมเปิลสเปซที่นับได้ (Finite sample space) ก่อให้เกิดเซตของจำนวนเลขจาก 0 ถึง 1 ซึ่งจำนวนเหล่านี้ สามารถนำไปใช้ประเมินผลความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้น

การหาความน่าจะเป็นเรากำหนดค่าของ "น้ำหนัก" ให้แก่ทุก ๆ จุดตัวอย่าง ซึ่งผลบวกของน้ำหนักเหล่านี้มีค่าเท่ากับ 1 ถ้าเรามีเหตุผลที่จะเชื่อได้ว่า จุดตัวอย่างอันหนึ่งอันใดมีโอกาสเกิดขึ้นอย่างแน่นอนในการทดลอง ค่าน้ำหนักของจุดตัวอย่างนั้นควรมีค่าเท่ากับ 1 ในกรณีตรงข้ามค่าน้ำหนักที่มีค่าใกล้ 0 จะถูกกำหนดให้สำหรับจุดตัวอย่างที่เกิดขึ้นได้ยาก หรือเกิดขึ้นน้อยมาก

ในการหาความน่าจะเป็นของเหตุการณ์ใด ๆ สมมุติ x เราบวกค่าน้ำหนักทั้งหมดของจุดตัวอย่างในเหตุการณ์ x นั้น ผลบวกเรียกว่าความน่าจะเป็นของเหตุการณ์ x และเขียนแทนด้วย Pr (x) ดังนั้นความน่าจะเป็นของเซตว่างเปล่าคือ Pr (o) จึงเท่ากับ 0 และความน่าจะเป็นของ Pr (S) เท่ากับ 1 ดังนั้น

$$0 < Pr(x) < 1, \quad Pr(o) = 0, \; Pr(S) = 1 \ldots\ldots\ldots\ldots\ldots (3.8)$$

ในการวิเคราะห์ความถี่ทางอุทกวิทยา ช่วงของการเกิดซ้ำหรือค่ารอบปีเป็นค่าช่วงระยะเวลาที่ค่าของเหตุการณ์ เท่ากับหรือมากกว่าค่าที่กำหนด จะเกิดขึ้น ค่ารอบปีแทนด้วยสัญลักษณ์ "T" มีหน่วยเป็น "ปี"

$$T = \frac{1}{Pr\ (x \geqslant x)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.9)$$

และ $Pr\ (x \geqslant x) = 1 - Pr\ (x \leqslant x) \ldots\ldots\ldots\ldots (3.10)$

ดังนั้น $$T = \frac{1}{1 - Pr\ (x \leqslant x)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.11)$$

3.3 การพิจารณาข้อมูลฝน

ในการวิเคราะห์ความถี่โดยวิธี Direct frequency analysis นั้น มีการเลือกใช้ข้อมูลอยู่ 2 แบบ คือ Annual series และ Partial duration series.

Annual series เป็นกลุ่มของข้อมูลฝนที่เลือกเอาเฉพาะค่าสูงสุดในแต่ละปี

Partial-duration series เป็นกลุ่มของข้อมูลที่เลือกเอาเฉพาะค่าที่มากกว่าค่าที่กำหนด

Annual series นั้นมีข้อเสียคือ ข้อมูลตัวที่ 2 หรือ 3 อาจจะมีค่ามากกว่าข้อมูลที่มีค่าสูงสุดของปีอื่น ๆ แต่กลุ่มข้อมูลแบบ Annual series นี้ ก็มักจะเป็นที่นิยมนำมาใช้ในการวิเคราะห์ ความถี่มากกว่า ทั้งนี้เพราะสะดวกในการเก็บข้อมูลและการวิเคราะห์แต่อย่างไรก็ตาม Partial-duration series ก็มีข้อดีที่สามารถหาข้อมูลในการวิเคราะห์ได้มากกว่า Annual series และจะเป็นผลดีสำหรับข้อมูลที่มีจำนวนปีน้อย ๆ

Kite ได้แสดงผลวิจัยของ Dalrymple ที่เปรียบเทียบค่ารอบปีของอนุกรมทั้งสองแบบ ดังตารางที่ แสดงให้เห็นว่าที่ค่ารอบปี 5 ปี มีค่ารอบปีต่างกัน 10 เปอร์เซ็นต์และที่ค่ารอบปี 10 ปี มีค่ารอบปีต่างกัน .5 เปอร์เซ็นต์

ตารางที่ 3.1 เปรียบเทียบค่ารอบปีของอนุกรมแบบ Partial-duration กับ Annual series (Kite)

รอบปี (ปี)	
Partial-duration	Annual series
.50	1.16
1	1.58
1.44	2
2	2.54
5	5.52
10	10.5
20	20.5
50	50.5
100	100.5

จากข้อพิจารณาดังที่ได้กล่าวมาแล้ว และเพื่อความสะดวกในการวิจัย ผู้วิจัยจึงได้เลือกใช้ข้อมูลแบบ Annual series มาใช้ในการวิจัยครั้งนี้

3.4 การวิเคราะห์ความถี่

ในการวิเคราะห์ความถี่ ของข้อมูลทางอุทกวิทยานั้นมีสมการทั่วไปและทฤษฎีที่นิยมใช้สำหรับการแจกแจงความถี่ดังนี้

3.4.1 สมการทั่วไปสำหรับการวิเคราะห์ความถี่

Chow (1964) ได้แสดงสมการทั่วไปสำหรับการวิเคราะห์ความถี่ดังนี้

$$x = \bar{x} + \Delta x \qquad \text{.........................(3.12)}$$

เมื่อ x คือ ค่าเหตุการณ์ใดเหตุการณ์หนึ่ง

$\bar{x}$ คือ ค่าเฉลี่ยและ Δx คือส่วนที่เบี่ยงเบนไปจากค่าเฉลี่ย ซึ่ง Δx นี้ขึ้นอยู่กับแนวโน้มที่จะเบี่ยงเบนไปจากค่าเฉลี่ย (ซึ่งอยู่ในรูปของความเบี่ยงเบนมาตรฐาน)

และความถี่ของเหตุการณ์ (ซึ่งอยู่ในรูปของ Frequency factor K)

ดังนั้น สมการทั่วไปสำหรับการวิเคราะห์ความถี่แสดงได้เป็น

$$x = \bar{x} + S.K. \qquad \text{.............(3.13)}$$

Chow (1964) ยังได้แสดงให้เห็นว่า สมการ (3.13) สามารถนำไปใช้กับทฤษฎีการแจกแจงความถี่แบบต่าง ๆ ได้ ซึ่งสมการ (3.13) แสดงได้เป็น

$$x_T = \bar{x}\,(1 + C_v.K) \qquad \text{............(3.14)}$$

เมื่อ x_T = ค่าของเหตุการณ์ที่รอบปีที่ T

$\bar{x}$ = ค่าเฉลี่ยของกลุ่มตัวอย่าง

S = ค่าความเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

K = ค่า Frequency factor

C_v = สัมประสิทธิ์ของความแปรปรวน

3.4.2 การแจกแจงแบบลอกนอร์มอลชนิด 2 พารามิเตอร์ (Two Parameter Lagnormal Distribution)

ถ้า ln x โดยที่ x เป็นตัวแปรผัน (variable) มีการแจกแจงแบบปกติ (Normally - distributed) ดังนั้นอนุกรมของตัวแปรผัน x จะมีการแจกแจงแบบลอกนอร์มอล (Logarithmic - normally distributed)

Kite (1977) ได้แสดงให้เห็นว่าการแจกแจงแบบลอกนอร์มอลชนิด 2 พารามิเตอร์มีค่า Probability density function เป็น

$$P(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} . e^{-\frac{[\ln x - \mu_y]^2}{2\sigma_y^2}} \qquad \text{..........(3.15)}$$

เมื่อ μ_y และ σ_y คือ ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของ lnx

และ Kite (1977) ยังได้แสดงสมการสำหรับหาค่า Frequency factor (K) ดังนี้

ถ้า $y = \ln x$ มีการแจกแจงแบบปกติ ดังนั้นสมการทั่วไปของความถี่สามารถเขียนได้ในรูปของลอก ดังนี้

$$\ln x_T = y_T = \mu_y + t\sigma_y \quad ....(3.16)$$

เมื่อ t เป็นค่าส่วนเบี่ยงเบนมาตรฐานปกติ (Standard normal deviate) หรือ t ก็คือค่า Frequency factor ของการแจกแจงแบบปกติ และค่า Frequency factor สำหรับการแจกแจงแบบลอกนอร์มอลชนิด 2 พารามิเตอร์แสดงได้ดังนี้

$$K = \frac{e^{\ln(1+C_v^2)^{\frac{1}{2}} t - \ln(1+C_v^2)/2} - 1}{C_v} \quad .....(3.17)$$

เมื่อ C_v เป็นสัมประสิทธิ์ของความแปรปรวน

3.4.3 การแจกแจงแบบลอกนอร์มอลชนิด 3 พารามิเตอร์ (Three-Parameter Lognormal Distribution)

เมื่อการแจกแจงแบบลอกนอร์มอล แทนการแจกแจงแบบปกติของลอก ตัวแปร x ดังนั้นลอกนอร์มอลชนิด 3 พารามิเตอร์ จะแทนการแจกแจงของลอกตัวแปร (x - a) เมื่อ a คือค่าขอบเขตล่าง (Lower boundary)

Kite (1977) ได้แสดงให้เห็นว่า การแจกแจงแบบลอกนอร์มอลชนิด 3 พารามิเตอร์ มีค่า probability density function เป็น

$$P(x) = \frac{1}{(x-a)\,\sigma_y\sqrt{2\pi}}\, e^{-\frac{[\ln(x-a) - \mu_y]^2}{2\sigma_y^2}} \quad ...(3.18)$$

เมื่อ μ_y และ σ_y^2 เป็นค่าเฉลี่ยและค่าความแปรปรวนของ $\ln(x-a)$ โดยที่ a เป็นค่าขอบเขตล่าง (Lower boundary) ซึ่งหาได้จากสมการ ดังนี้

ถ้า Z_1 และ Z_2 เป็นค่าสัมประสิทธิ์ความแปรปรวนของการแจกแจงของตัวแปร x และ x - a ดังนั้น

$$Z_1 = \sigma/\mu \quad ....(3.19)$$

และ

$$Z_2 = \sigma/_{(\mu - a)} \quad ....(3.20)$$

จากสมการ (3.19) และ (3.20) จะได้ค่า a ดังนี้

$$a = \mu - \frac{\sigma}{Z_2} \quad ....(3.21)$$

ค่า Z_2 สามารถหาได้จากสมการ

$$\gamma_1 = 3Z_2 + Z_2^3 \quad ....(3.22)$$

เมื่อ γ_1 เป็นค่าสัมประสิทธิ์ของความเบ้ของการแจกแจงของตัวแปร x (Coefficient of skew of the distribution)

สมการ (3.22) เขียนได้เป็น

$$Z_2 = \frac{1 - \omega^{2/3}}{\omega^{1/3}} \quad ....(3.23)$$

เมื่อ

$$= \frac{-\gamma_1 + (\gamma_1^2 + 4)^{1/2}}{2} \quad ....(3.24)$$

ดังนั้น ค่า μ_y และ σ_y สามารถหาได้จากสมการดังนี้

$$\mu_y = \ln(\sigma/Z_2) - \frac{1}{2}\ln(Z_2^2 + 1) \quad .....(3.25)$$

$$\sigma_y = \left[\ln(Z_2^2 + 1)\right]^{\frac{1}{2}} \quad ....(3.26)$$

Kite (1977) ได้แสดงสมการสำหรับหาค่า Frequency factor ของการแจกแจงแบบลอกนอร์มอลชนิด 3 พารามิเตอร์ ดังนี้

$$K = \frac{\exp\{[\ln(1+Z_2^2)]^{\frac{1}{2}}t - [\ln(1+Z_2^2)]/2\} - 1.0}{Z_2} \quad ...(3.27)$$

เมื่อ t คือค่า Frequency factor ของการแจกแจงแบบปกติ

3.4.4 การแจกแจงแบบกัมเบล (Gumbel's Distribution)

Kite (1977) ได้แสดงให้เห็นว่าการแจกแจงแบบกัมเบล มีความน่าจะเป็นสะสม (Cumulative probability) เป็น

$$P(x) = e^{-e^{-\alpha(x-\beta)}} \qquad \text{................(3.28)}$$

และมีค่า Probability density เป็น

$$P(x) = \alpha\, e^{\{-\alpha(x-\beta)-e^{-\alpha(x-\beta)}\}} \qquad \text{.........(3.29)}$$

โดยที่

α คือ Concentration parameter และ

β คือ การจัดแนวโน้มเข้าสู่ส่วนกลาง (Measure of central tendency)

และได้ประมาณค่า Parameter ไว้ดังนี้

$$\alpha = 1.2825/S \qquad \text{..............(3.30)}$$

$$\beta = \bar{x} - 0.4500\ S \qquad \text{..............(3.31)}$$

โดยที่

$\bar{x}$ คือ ค่าเฉลี่ยของตัวอย่าง

S คือ ค่าความเบี่ยงเบนมาตรฐานของตัวอย่าง

Kite (1977) ยังได้เสนอสมการสำหรับการหาค่า Frequency factor ของการแจกแจงแบบกัมเบลดังนี้

จากการแจกแจงของความน่าจะเป็นสะสม (Cumulative probability) จะได้

$$Y_T = -\ln(-\ln(T-1)/T) \qquad \text{...........(3.32)}$$

โดยที่ Y_T เป็น reduced variable

และถ้า

n คือ จำนวนค่าของเหตุการณ์เรียงลำดับจากมากไปหาน้อย

m คือ ลำดับค่าเหตุการณ์ ดังนั้น m = 1 สำหรับค่าของเหตุการณ์มากที่สุด และ m = n สำหรับค่าของเหตุการณ์น้อยที่สุด

ดังนั้น $T = \frac{(n + 1)}{m}$ และสมการ (3.32) สามารถเขียนได้เป็น

$$Y_m = -\ln\left| -\ln \{(n+1-m) / (n+1)\} \right| \qquad \text{..........(3.33)}$$

ถ้าค่าเฉลี่ย $\bar{Y}_n$ และค่าความเบี่ยงเบนมาตรฐาน S_n ของอนุกรม Y_m เมื่อ $m = 1,2,...n$ คำนวณได้จาก

$$\bar{Y}_n = \sum_{i=1}^{n} Y_{m/n} \qquad \text{..............(3.34)}$$

และ

$$S_n = \sqrt{\frac{\sum_{m=1}^{n} (Y_m - \bar{Y}_n)^2}{n}} \qquad \text{.............(3.35)}$$

ค่า Frequency factor ของการแจกแจงแบบกัมเบลหาได้ดังนี้

$$K = \frac{Y_m - \bar{Y}_n}{S_n} \qquad \text{..............(3.36)}$$

3.4.5 การแจกแจงแบบเพียร์สันแบบที่ 3 (Pearson Type III Distribution)

Kite (1977) แสดงค่า Probability density function ของการแจกแจงแบบเพียร์สันแบบที่ 3 ดังนี้

$$P(x) = \frac{1}{\alpha \, \Gamma(\beta)} \left\{\frac{x - \gamma}{\alpha}\right\}^{\beta - 1} e^{-\left\{\frac{x-\gamma}{\alpha}\right\}} \qquad \text{.....(3.37)}$$

เมื่อ α β และ γ เป็นพารามิเตอร์ และ $\Gamma(\beta)$ คือค่า Gamma function

แทนค่า $y = (x - \gamma)/\alpha$ สมการ (3.37) เขียนได้เป็น

$$P(y) = \frac{y^{\beta-1} e^{-y}}{\Gamma(\beta)} \qquad \text{.................(3.38)}$$

ค่า พารามิเตอร์ α β และ γ หาได้จากสมการดังนี้

$$\alpha = \sigma / \sqrt{\beta} \qquad \text{...............(3.39)}$$

$$\beta = (2/\gamma_1)^2 \qquad \text{...............(3.40)}$$

และ

$$\gamma = \mu - \sigma\sqrt{\beta} \qquad \text{..............(3.41)}$$

โดยที่ μ คือ ค่าเฉลี่ย (mean) ของตัวอย่าง

σ คือ ค่าความเบี่ยงเบนมาตรฐาน (Standard deviation) ของตัวอย่าง

γ_1 คือ สัมประสิทธิ์ของความเบ้ (Coefficient of skew)

ค่า Frquency factor ของการแจกแจงแบบเพียร์สันแบบที่ 3 Kite (1977) แสดงไว้ดังนี้

$$K = t + (t^2-1)\frac{\gamma_1}{6} + \frac{1}{3}(t^3 - t)\left(\frac{\gamma_1}{6}\right)^2 - (t^2-1)\left(\frac{\gamma_1}{6}\right)^3 + t\left(\frac{\gamma_1}{6}\right)^4 + \frac{1}{3}\left(\frac{\gamma_1}{6}\right)^5 \qquad \text{.........(3.42)}$$

โดยที่ t เป็นค่า Frequency factor ของการแจกแจงแบบปกติ

3.4.6 การแจกแจงแบบลอก-เพียร์สันแบบที่ 3 (Log-Pearson Type III Distribution)

การแจกแจงแบบนี้ U.S. Federal Water Resources Council เสนอต่อรัฐบาลสหรัฐอเมริกาเมื่อปี 1967 เพื่อใช้เป็นวิธีมาตรฐานสำหรับการวิเคราะห์ความถี่น้ำหลาก

Kite (1977) ได้เสนอไว้ดังนี้

ถ้า lnx ของตัวแปร x ถูกแจกแจง แบบ เพียร์สันแบบที่ 3 ดังนั้นตัวแปร x ก็จะแจกแจงแบบ ลอกเพียร์สันแบบที่ 3 และมี Probability function เป็น

$$P(x) = \frac{1}{\alpha\, x\, \Gamma(\beta)} \left\{\frac{\ln x - \gamma}{\alpha}\right\}^{\beta-1} e^{-\left\{\frac{\ln x - \gamma}{\alpha}\right\}} \qquad \text{......(3.43)}$$

เพื่อ α, β และ γ คือ scale, shape และ location parameters ตามลำดับ

$\Gamma(\beta)$ คือ Gamma function

ค่า Frequency factor สำหรับการแจกแจงแบบลอกเพียร์สันแบบที่ 3 สามารถได้จาก สมการที่ (3.42), Kite (1977) แต่จะต้องใช้ค่า γ_y ซึ่งเป็นสัมประสิทธิ์ความเบ้ของเหตุการณ์ (Coefficient of skew of the events)

3.4.7 ค่าความคลาดเคลื่อนมาตรฐาน (Standard error)

หมายถึงค่าความคลาดเคลื่อนของค่าเหตุการณ์ที่ประเมินได้ และจากการวิเคราะห์ความถี่โดยใช้ทฤษฎีการแจกแจงแบบต่าง ๆ ดังที่ได้กล่าวมาในหัวข้อ 3.4.1 ถึง 3.4.7 Kite (1977) ได้แสดงสมการสำหรับการหาค่า ความคลาดเคลื่อนมาตรฐานของการแจกแจงแบบต่าง ๆ และในที่นี้จะขอกล่าวแต่เพียงค่าความคลาดเคลื่อนมาตรฐานของการแจกแจงแบบกัมเบลเท่านั้น ซึ่งจะนำมาใช้ในการวิจัยครั้งนี้

Lowery และ Nash (Kite 1977) ได้เสนอสมการสำหรับหาค่าความคลาดเคลื่อนมาตรฐานของทฤษฎีการแจกแจงแบบกัมเบลดังนี้

$$S.E. = \frac{S^2}{n} (1 + 1.1396 K + 1.1 K^2) \quad \text{........(3.44)}$$

โดยที่

S.E. = ค่าความคลาดเคลื่อนมาตรฐาน (Standard error)

S = ค่าความเบี่ยงเบนมาตรฐาน (Standard deviation)

n = จำนวนข้อมูล

K = ค่า Frequency factor สำหรับการแจกแจงแบบกัมเบล

3.5 การเลือกทฤษฎีการแจกแจงความถี่

การเลือกใช้ทฤษฎีการแจกแจงความถี่ที่เหมาะสมกับภูมิภาค จะทำให้การวิเคราะห์ได้ผลดี ฉะนั้นการที่จะเลือกใช้ทฤษฎีการแจกแจงแบบไหนนั้นน่าที่จะต้องทำการวิเคราะห์โดยใช้ทฤษฎีการแจกแจงหลาย ๆ แบบแล้วทำการทดสอบหาความเหมาะสมของทฤษฎีนั้น ๆ ว่าทฤษฎีการแจกแจงแบบไหนเหมาะสมกว่า

สำหรับประเทศไทย Ertuna (1970), Bhuiyan (1982) ได้วิจัยเปรียบเทียบการแจกแจงความถี่แบบต่าง ๆ ว่าวิธีใดเหมาะสมดีสำหรับประเทศไทย และทั้งสองได้สรุปตรงกันว่าทฤษฎีการแจกแจงแบบกัมเบลเหมาะสมดีสำหรับประเทศไทย (ยกเว้นภาคใต้ของประเทศ) มากกว่าทฤษฎีการแจกแจงแบบอื่น ฉะนั้นการวิจัยครั้งนี้ผู้วิจัยจึงเลือกใช้ทฤษฎีการแจกแจงแบบกัมเบล

3.6 ความถูกต้องของผลวิเคราะห์ความถี่

ความถูกต้องของผลวิเคราะห์ความถี่ที่ประเมินได้จาก

3.6.1 จำนวนปีที่มีการบันทึกข้อมูลไว้อย่างต่อเนื่อง กล่าวคือถ้ามีจำนวนปีที่บันทึกข้อมูลไว้มากก็จะทำให้มีความมั่นใจในการประเมินได้มาก และค่าประเมินของรอบปีสั้น ๆ จะมีค่า Standard error น้อยแต่ถ้ามีค่ารอบปีมาก ก็จะมีค่า Standard error มาก ถ้ามีการบันทึกข้อมูลไว้น้อย

3.6.2 ทฤษฎีการแจกแจงที่ใช้วิเคราะห์

ทฤษฎีที่เหมาะสมที่สุด จะได้ผลวิเคราะห์ที่ดีที่สุด

3.6.3 ความถูกต้องของข้อมูล

ข้อมูลหากไม่ถูกต้องก็จะทำให้ผลการวิเคราะห์ผิดพลาดได้ ในการอ่านค่าของข้อมูลปริมาณฝน มักจะมีความคลาดเคลื่อนไม่มากนัก สำหรับฝนที่มีช่วงเวลายาว แต่ค่าปริมาณฝนที่มีช่วงเวลาสั้นนั้น มักจะมีความคลาดเคลื่อน แต่ทั้งนี้ก็ขึ้นอยู่กับประสิทธิภาพของเครื่องมือ การติดตั้งกราฟและผู้อ่านกราฟฝนว่ามีความละเอียดเพียงใด

3.7 หลักการโดยทั่วไปของความสัมพันธ์ระหว่างปริมาณฝน-ช่วงเวลา-ความถี่ (Generalized rainfall-duration-frequency relationships)

หลักการโดยทั่วไปของความสัมพันธ์ของฝนในการวิจัยครั้งนี้ จะทำการหาความสัมพันธ์ของฝนที่มีช่วงเวลาตั้งแต่ 15 นาที จนถึง 24 ชั่วโมง ทั้งนี้เพื่อที่จะได้นำหลักการของความสัมพันธ์นี้ไปประยุกต์ใช้ในการประเมินค่าปริมาณฝนที่มีช่วงเวลาสั้น ๆ ในบริเวณที่ขาดแคลนข้อมูลฝน

3.7.1 อัตราส่วนปริมาณฝน-ช่วงเวลา (Depth-Duration Ratios)

หมายถึงค่าอัตราส่วนของค่าปริมาณฝน-ช่วงเวลาที่ช่วงเวลาใด ๆ หมายถึงอัตราส่วนของค่าปริมาณฝนที่มีช่วงเวลานั้นต่อค่าปริมาณฝนที่มีช่วงเวลาที่กำหนดในรอบปีเดียวกัน

จากผลการวิจัยของ U.S. Weather Bureau พบว่าอัตราส่วนนี้ไม่ต่างกันมากนักในแต่ละภูมิภาคของอเมริกา และต่อมา Reich (1963) ยังพบว่าในอัฟริกาใต้กับสหรัฐอเมริกายังคงมีค่าใกล้เคียงกัน จึงเสนอแนะว่าอาจจะนำไปใช้ในภูมิภาคอื่นของโลกที่ขาดแคลนข้อมูลฝนได้

Bell (1969) ได้วิจัยหาความสัมพันธ์ของปริมาณฝน-ช่วงเวลาในสหรัฐอเมริกาและออสเตรเลีย แล้วเสนอว่า ค่าอัตราส่วนปริมาณฝน-ช่วงเวลา (เมื่อเอาค่าปริมาณฝนที่มีช่วงเวลา 1 ชั่วโมงเป็นหลักที่ช่วงเวลาใด ๆ ตั้งแต่ 5 นาทีถึง 2 ชั่วโมง ที่ได้จากสถานีฝนในประเทศต่าง ๆ จะมีค่าใกล้เคียงกัน

สวามี หอสุชาติ (1983) ได้วิจัยหาความสัมพันธ์ของประมาณฝน-ช่วงเวลาในภาคเหนือของประเทศไทย และได้เปรียบเทียบกับผลการวิจัยของ Bell (1969) พบว่าค่าอัตราส่วนปริมาณฝน-ช่วงเวลาที่ช่วงเวลา 15 นาที ถึง 2 ชั่วโมง ที่ได้จากสถานีฝนในภาคเหนือของประเทศไทย มีค่าใกล้เคียงกับค่าที่ได้จาก สหรัฐอเมริกาและออสเตรเลีย ดังที่แสดงไว้ในตารางที่ 3.2

ตารางที่ 3.2 เปรียบเทียบค่าอัตราส่วนปริมาณฝน-ช่วงเวลาระหว่างสหรัฐอเมริกา ออสเตรเลีย และภาคเหนือของประเทศไทย (Bell (1969) และสวามี หอสุชาติ (1983))

พื้นที่	ช่องเวลา			
	5 นาที	15 นาที	30 นาที	2 ชั่วโมง
สหรัฐอเมริกา				
ค่าอัตราส่วนเฉลี่ย	0.29	0.57	0.79	1.25
ความเบี่ยงเบนมาตรฐาน	0.03	0.04	0.04	0.08
ออสเตรเลีย				
ค่าอัตราส่วนเฉลี่ย	0.30	0.57	0.78	1.24
ความเบี่ยงเบนมาตรฐาน	0.04	0.04	0.03	0.06
ภาคเหนือของประเทศไทย				
ค่าอัตราส่วนเฉลี่ย	0.193	0.478	0.732	1.247
ความเบี่ยงเบนมาตรฐาน	0.024	0.074	0.075	0.031

3.7.2 อัตราส่วนปริมาณฝน-ความถี่ (Depth-frequency Ratios)

ค่าอัตราส่วนปริมาณฝน-ความถี่ที่ค่ารอบปี T ปี หมายถึงค่าอัตราส่วนของปริมาณฝนในรอบ T ปี ต่อปริมาณฝนในรอบปีที่กำหนด ที่ช่วงเวลาเดียวกัน

Stewart (1960) ได้เสนอว่า ค่าอัตราส่วนนี้ของแต่ละประเทศจะมีค่าใกล้เคียงกัน

สวามี หอสุชาติ (1983) ได้วิจัย ความสัมพันธ์ของปริมาณฝน-ความถี่ในภาคเหนือของประเทศไทย และได้เปรียบเทียบกับผลการวิจัยของ Bell (1969) ซึ่งได้วิจัยหาความสัมพันธ์ของปริมาณฝน-ความถี่ในประเทศสหรัฐอเมริกา และออสเตรเลีย พบว่าค่าปริมาณฝน-ความถี่ (โดยของค่ารอบปี 10 ปี เป็นหลัก) พบว่ามีค่าใกล้เคียงกันดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 เปรียบเทียบค่าอัตราส่วนปริมาณฝน-ความถี่ (โดยเอาค่าปริมาณฝนในรอบ 10 ปี เป็นหลัก) ระหว่างสหรัฐอเมริกา, ออสเตรเลีย และภาคเหนือของประเทศไทย Bell (1969), สวามี หอสุชาติ (1983)

รอบปี	สหรัฐอเมริกา		ออสเตรเลีย		ภาคเหนือของประเทศไทย	
	อัตราส่วนเฉลี่ย	ความเบี่ยงเบนมาตรฐาน	อัตราส่วนเฉลี่ย	ความเบี่ยงเบนมาตรฐาน	อัตราส่วนเฉลี่ย	ความเบี่ยงเบนมาตรฐาน
1	0.54	0.05	0.52	0.05	-	-
2	0.63	0.05	0.65	0.05	0.653	0.054
5	0.85	0.03	0.85	0.05	0.862	0.022
25	1.17	0.05	1.18	0.03	1.174	0.027
50	1.31	0.06	1.33	0.06	1.304	0.047
100	1.46	0.07	1.50	0.08	1.432	0.067

3.8 การประเมินค่าปริมาณฝนที่มีช่วงเวลาสั้น ๆ ในบริเวณที่ขาดแคลนข้อมูลฝน

ได้มีผู้ศึกษาหาวิธีการประเมินค่าปริมาณฝนที่มีช่วงเวลาสั้น ๆ ในบริเวณที่ขาดแคลนข้อมูลฝน ซึ่งแต่ละคนก็ใช้วิธีแตกต่างกันไป และในการประเมินค่าปริมาณฝนส่วนใหญ่ จะใช้ข้อมูลฝนรายวันมาเกี่ยวข้องด้วย รายละเอียดการประเมินค่าปริมาณฝนที่ได้เสนอมาแล้วดังกล่าว จากการวิจัยของ สวามี หอสุชาติ (1983) ก็ได้กล่าวถึงวิธีการของแต่ละบุคคลแล้ว ผู้วิจัยจึงไม่ขอนำมากล่าวอีก

สำหรับการวิจัยครั้งนี้จะได้เสนอวิธีการประเมินค่าปริมาณฝนที่มีช่วงเวลาสั้น ๆ ในบริเวณที่ขาดแคลนข้อมูลฝน ในรูปของความสัมพันธ์ปริมาณฝน-ช่วงเวลา, ปริมาณฝน-ความถี่ และแผนที่เส้นชั้นระดับปริมาณฝนเท่ากัน และแผนที่เส้นชั้นค่าอัตราส่วนปริมาณฝน-ช่วงเวลาเท่ากัน

3.9 การเสนอผลการวิจัยเพื่อการประเมินค่าปริมาณฝนในภูมิภาคที่ต้องการ

จากการวิจัยที่ผ่านมาได้มีผู้วิจัยเสนอผลการวิจัยใช้ในการประเมินค่าปริมาณฝนในภูมิภาคที่ต้องการ ซึ่งสรุปได้ดังนี้

3.9.1 กราฟแสดงความสัมพันธ์ของความเข้มฝน-ช่วงเวลา-ความถี่ ของแต่ละสถานีฝนในภูมิภาคนั้น เช่น การวิจัยของ ธำรง เปรมปรีดิ์ (2521) และ Mustonen (1969)

3.9.2 สมการของความสัมพันธ์ที่ได้จากการทดลอง (Empirical equation) กับแผนที่แสดงค่าคงที่เท่ากัน หรือไดอาแกรมที่แสดงค่าพารามิเตอร์ที่จะใช้ในแต่ละภูมิภาค เช่น การวิจัยของ Lambor (1967) และ Australia, Institution of Engineers (1958)

3.9.3 แผนที่แสดงค่าปริมาณฝนเท่ากัน (ในรอบปีและช่วงเวลาที่กำหนด) เส้นชั้นค่าอัตราส่วนปริมาณฝน-ช่วงเวลาเท่ากัน เช่นการวิจัยของ สวามี หอสุชาติ (1983), Bell (1964) Reich (1963) และ Hershfield

3.10 เส้นชั้นระดับปริมาณน้ำฝนเท่ากัน

ในการวิจัยครั้งนี้ จะได้เสนอผลการวิจัยในรูปของแผนที่แสดงค่าปริมาณฝนเท่ากัน (ที่ช่วงเวลาและรอบปีที่กำหนด) และค่าอัตราส่วนเท่ากันของค่าปริมาณฝน-ช่วงเวลา

การเสนอผลการวิจัยในรูปของเส้นชั้นฝนที่เท่ากันนี้ จะให้ผลที่ถูกต้อง หรือใกล้เคียงกับความเป็นจริงแค่ไหนนั้น ย่อมขึ้นกับค่าความหนาแน่นของสถานีฝนที่สามารถให้ข้อมูลฝนว่ามีการกระจายอย่างไรบ้าง

การเขียนเส้นชั้นระดับปริมาณฝนเท่ากัน หรือค่าอัตราส่วนปริมาณฝนเท่ากัน มีหลักเกณฑ์พอสรุปได้ดังนี้

3.10.1 กำหนดค่าของผลวิเคราะห์ลงไปบนแต่ละสถานีฝน

3.10.2 การเขียนเส้นชั้นระดับให้ใช้วิธีการเทียบสัดส่วนโดยตรงระหว่างค่าที่ได้จากสถานีฝนที่อยู่ใกล้ที่สุด

3.10.3 ช่วงชั้นระดับควรเลือกให้เหมาะสมไม่ถี่หรือห่างเกินไป

3.10.4 การเขียนเส้นชั้นระดับ จะต้องลากเส้นด้วยมือ ลักษณะเส้นจะต้องโค้งเรียบ

3.11 หลักเกณฑ์ในการเลือกค่ารอบปีเพื่อใช้ในการออกแบบ

หลักเกณฑ์ที่ใช้คัดเลือกค่ารอบปีที่เหมาะสมเพื่อใช้ในการออกแบบนั้น Nemec และ Camp, Dresser & Mc Kee เสนอแนะไว้ดังนี้

ตารางที่ 3.4 การเลือกค่ารอบปีเพื่อการออกแบบ

งาน	รอบปี	เปอร์เซ็นต์ที่อาจเกิดสูงกว่า
1. ทางระบายน้ำ, ท่อลอด ที่ไม่สำคัญ	2	50
2. ทางระบายน้ำ ท่อลอด การป้องกันน้ำท่วมจากสถาบันต่าง ๆ เช่นโรงพยาบาล โรงเรียน สถานที่ราชการ	5	20
3. ฝายน้ำล้น,งานกักเก็บน้ำขนาดเล็กตามหมู่บ้านในชนบท	10-20	10-5
4. ท่อลอด ทางสาธารณะระหว่างหมู่บ้าน ฝายน้ำล้นขนาดเล็ก การระบายน้ำจากโรงงานอุตสาหกรรม	30-50	3-2
5. ฝายน้ำล้นหรือการระบายน้ำจากบริเวณที่อาจเกิดการเสียหายถึงชีวิต	50-100	2-1

3.12 การหาปริมาณน้ำฝนที่บริเวณใด ๆ

การหาปริมาณน้ำฝนที่บริเวณใดสามารถหาได้ดังนี้

1. หา Concentration time (T) ซึ่งเป็นระยะเวลาในการเดินทางของน้ำจากจุดไกลสุด มายังจุดนำน้ำทิ้งออกไปนอกบริเวณ

2. หาค่าช่วงเวลาการตกของฝน (D) ซึ่งถ้า D=T จะให้การไหลสูงสุด ดังนั้นค่าช่วงเวลาการตกของฝนซึ่งเท่ากับค่า Concentration time

3. หาอัตราการตกสูงสุดของฝนที่รอบปีที่ต้องการ (I)

4. หาปริมาณน้ำฝนได้จากผลคูณของ D กับ I

3.13

การหาค่า Frequency factor ที่มีค่ารอบปี T ต่าง ๆ กัน สำหรับการแจกแจงแบบกัมเบล สามารถหาได้ดังนี้

1. หาค่า Y_T จากสมการ (3.32)

2. เมื่อทราบว่า n และ m ของอนุกรมข้อมูลฝน หาค่า Y_m จากสมการ (3.33)

3. หาค่าเฉลี่ย $(\bar{Y}_m)$ ของอนุกรม Y_m จากสมการ (3.34)

4. หาค่า ความเบี่ยงเบนมาตรฐานของอนุกรม Y_m จากสมการ (3.35)

5. หาค่า Frequency factor (K) ที่รอบปี T จากสมการ (3.36)