บทที่ 3

ทฤษฎีของความน่าจะเป็นและสถิติ

ขบวนการทางอุทกวิทยา (hydrologic process) มีความยุ่งยากสลับซับซ้อนมากมีปรากฏการณ์ธรรมชาติทางอุทกวิทยาเป็นจำนวนมาก ที่ไม่สามารถจะหาคำตอบที่แน่นอนได้โดยตรง เพียงแต่สามารถที่จะแปรความหมายหรืออธิบายเกี่ยวกับปรากฏการณ์นั้น ๆ ในแนวความคิดของความน่าจะเป็นเท่านั้น ทั้งนี้เพราะสาเหตุของการเกิดเหตุการณ์ทางอุทกนั้นมีความไม่แน่นอน ขึ้นกับตัวแปรต่าง ๆ ที่เป็นปรากฏการณ์ธรรมชาติที่ไม่สามารถจะกำหนดหรือควบคุมได้

การเกิดฝน (rainfall) เป็นขบวนการทางอุทกอย่างหนึ่ง ที่ต้องอาศัยการแปรความหรืออธิบายเกี่ยวกับขนาดของเหตุการณ์ (magnitude of event) ที่จะเกิดขึ้นในลักษณะของความน่าจะเป็น จึงได้มีผู้พยายามคิดหาวิธีการทางสถิติ (statistical method) เพื่อใช้ในการวิเคราะห์ข้อมูลฝน ในปี ค.ศ. 1935 Yarnell ได้คิดทฤษฎีเพื่อใช้สำหรับวิเคราะห์ความถี่ (frequency) ของข้อมูลฝนในสหรัฐอเมริกาขึ้น และได้เสนอเพื่อใช้งานจนเป็นที่รู้จักกันอย่างแพร่หลายในชื่อของทฤษฎีวิเคราะห์ความถี่ข้อมูลฝนคือ "Yarnell rainfall frequency data" และต่อมาได้มีผู้ประยุกต์ใช้ทฤษฎีของความน่าจะเป็นแบบต่าง ๆ ที่นิยมใช้สำหรับวิเคราะห์ข้อมูลทางอุทกวิทยามาใช้ในการวิเคราะห์ข้อมูลฝน อีกหลายทฤษฎีด้วยกัน ปัจจุบันก็ยังไม่สามารถที่จะหาวิธีการอื่นใดมาใช้ทำนายเกี่ยวกับอัตราการตกสูงสุดที่ช่วงเวลาต่าง ๆ ของแต่ละปีได้เป็นที่แน่นอน ยังคงต้องอาศัยวิธีการทำนายด้วยหลักการของความน่าจะเป็นและสถิติ สำหรับหลักการของความน่าจะเป็นและสถิติที่เกี่ยวข้องนั้นจะแสดงไว้โดยสังเขปต่อไป

3.1 <u>หลักการเกี่ยวกับความน่าจะเป็นและสถิติที่เกี่ยวข้อง</u>

3.1.1 ตัวแปรทางสถิติ (Statistical Variables)

เมื่อนำหลักการทางสถิติมาประยุกต์ใช้กับข้อมูลทางอุทกวิทยานั้น สามารถจะอธิบายหลักการเปรียบเทียบเกี่ยวกับตัวแปรทางด้านสถิติกับข้อมูลทางอุทกวิทยาได้ว่า ในทางสถิตินั้นค่าสังเกต (objects) ทั้งหมดที่เกี่ยวข้องในการทดลองหรือการสำรวจ หมายถึงประชากร (population) ส่วนต่าง ๆ ที่เปลี่ยนแปลงค่าได้ซึ่งประกอบขึ้นเป็นประชากรนั้นเรียกว่าตัวแปร (variables) ซึ่งโดยทั่วไปมักจะแทนด้วยอักษร X ค่าของตัวแปร (variate) แต่ละตัวจะแทน

ด้วยอักษร x สำหรับปรากฏการณ์ทางอุทก (hydrologic phenomena) เมื่อให้ตัวแปร X เป็นปริมาณฝนที่ตกในช่วงเวลา (duration) ตามกำหนดที่ได้จากการตรวจวัดแต่ละครั้ง x ก็จะเป็นค่าของปริมาณฝนตามที่ทำการตรวจวัด เช่นปริมาณฝนรายวัน (daily rainfall) ที่ได้จากการตรวจวัดครั้งหนึ่งเท่ากับ 142.7 ม.ม. X คือ ปริมาณฝนรายวัน และ x คือค่าปริมาณฝนที่วัดได้เท่ากับ 142.7 ม.ม. เป็นต้น

การทดลองใด ๆ ที่ไม่สามารถจะทราบผลล่วงหน้าได้ เรียกว่าการทดลองสุ่ม (random trial) ผลลัพธ์ที่ได้จากการทดลองสุ่มในแต่ละครั้ง เรียกว่าตัวแปรสุ่ม (random variable) ผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองอย่างหนึ่ง เรียกว่า แซมเปิลสเปซ (sample space) ในทางอุทกนั้น การสังเกต (observation) ปรากฏการณ์ใด ๆ ที่มีระยะเวลาแน่นอน คือการทดลองสุ่ม เช่นการวัดค่าปริมาณฝนที่ตกในแต่ละวัน ปริมาณฝนที่ตกในแต่ละวัน เรียกว่าตัวแปรสุ่ม สำหรับค่าปริมาณฝนนั้นจะมีค่าเป็นจำนวนเต็ม และมีแซมเปิลเปซแบบนับไม่ได้ (infinite)

3.1.2 พารามิเตอร์ทางสถิติ (Statistical Parameters)

พารามิเตอร์ทางสถิตินั้น จะเป็นตัวบอกถึงลักษณะเฉพาะ (characteristic) ของการแจกแจงของสถิติ (statistical distribution) พารามิเตอร์ทางสถิตินั้นมีอยู่เป็นจำนวนมาก แต่ในที่นี้จะกล่าวถึงเฉพาะตัวที่สำคัญและเกี่ยวข้องพอสังเขป คือ

ก. การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measure of Central Tendency)

พารามิเตอร์โดยทั่วไปที่แทนการวัดแนวโน้มเข้าสู่ส่วนกลางของการแจกแจงเกี่ยวกับสถิติ ก็คือค่าเฉลี่ยที่สำคัญ ๆ ได้แก่

(1) ค่าเฉลี่ยเลขคณิต (arithmatic mean) ถ้า X เป็นตัวแปรสุ่ม และ $x_1, x_2, \ldots, x_n$ เป็นค่าของตัวแปรสุ่มที่ได้จากการทดลอง n ครั้ง ค่าเฉลี่ยของตัวเลขเหล่านี้ เรียกว่าค่าเฉลี่ยเลขคณิต เขียนแทนด้วยสัญญลักษณ์ $\bar{x}$ ซึ่งสามารถหาค่าได้จากสมการต่อไปนี้

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \ldots\ldots\ldots(3\text{-}1)$$

เมื่อ X คือตัวแปรสุ่มของตัวอย่าง

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \text{............(3-2)}$$

เมื่อ X คือตัวแปรสุ่มของประชากร

(2) มัธยฐาน (median) หมายถึงค่าของตัวแปรสุ่มที่แบ่งการแจกแจงเป็น 2 ส่วน โดยแต่ละส่วนมีความน่าจะเป็นเท่ากับ 0.5 ถ้าเขียนกราฟของการแจกแจงแล้ว มัธยฐานจะเป็นจุดแบ่งพื้นที่ใต้โค้งออกเป็น 2 ส่วนเท่า ๆ กัน ซึ่งถ้ากล่าวถึงค่ามัธยฐานของกลุ่มตัวอย่างหรือประชากรแล้ว ก็คือค่ากึ่งกลางของข้อมูลจำนวนคี่และค่าเฉลี่ยกึ่งกลางของ 2 ค่า สำหรับข้อมูลจำนวนคู่

(3) ฐานนิยม (mode) เป็นค่าของตัวแปรสุ่มที่มีค่าของฟังชั่นของความน่าจะเป็นสูงสุด ในข้อมูลบางชุดไม่มีค่าฐานนิยมหรืออาจจะมีค่าฐานนิยมมากกว่า 1 ค่าได้

ข. การวัดการกระจาย (Measure of Variability)

สำหรับข้อมูลต่างชุดกัน ที่มีค่าเฉลี่ยเท่ากันนั้น อาจจะมีลักษณะการแจกแจงต่างกัน และวิธีการที่จะบอกว่าข้อมูลชุดใดมีการกระจายมากน้อยกว่ากันเพียงใดนั้น มีวิธีการที่จะบอกได้ด้วยค่าต่อไปนี้

(1) ความแปรปรวน (Variance) เป็นค่าที่จะบอกให้ทราบว่าค่าของตัวแปรสุ่มห่างจากค่าเฉลี่ยมากน้อยเพียงใด เมื่อใช้ S^2 และ σ^2 เป็นสัญลักษณ์แทนค่าความแปรปรวนของตัวอย่าง และประชากรตามลำดับแล้ว สามารถหาค่าความแปรปรวนได้จากสมการ

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} \left[(x_i - \bar{x})^2 \right] \qquad \text{..........(3-3)}$$

สำหรับตัวแปรสุ่มที่เป็นตัวอย่างขนาด n จำนวน

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\mu)^2\right] \qquad \text{..........} (3\text{-}4)$$

สำหรับตัวแปรสุ่มที่เป็นประชากรขนาด n จำนวน

(2) ความเบี่ยงเบนเฉลี่ย (mean deviation) เป็นค่าของตัวแปรสุ่ม ที่เบี่ยงเบนไปจากค่าเฉลี่ย โดยไม่คิดเครื่องหมายแทนด้วยสัญญลักษณ์ M.D. สามารถหาค่าได้จากสมการ

$$\text{m.d.} = \frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}| \qquad \text{..........} (3\text{-}5)$$

สำหรับตัวแปรสุ่มที่เป็นตัวอย่างขนาด n จำนวน

$$\text{M.D.} = \frac{1}{n}\sum_{i=1}^{n}|x_i-\mu| \qquad \text{..........} (3\text{-}6)$$

สำหรับตัวแปรสุ่มที่เป็นประชากรขนาด n จำนวน

(3) ความเบี่ยงเบนมาตรฐาน (standard deviation) คือค่ารากที่สองของความแปรปรวนของตัวอย่างและประชากรแทนด้วยสัญญลักษณ์ S และ σ ตามลำดับ

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{n}{n-1}\left(\overline{x^2}-\bar{x}^2\right)} \qquad \text{.....} (3\text{-}7)$$

สำหรับตัวแปรสุ่มที่เป็นตัวอย่างขนาด n จำนวน

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}} = \sqrt{\frac{n}{n-1}\left(\overline{x^2}-\mu^2\right)} \qquad \text{.....} (3\text{-}8)$$

สำหรับตัวแปรสุ่มที่เป็นประชากรขนาด n จำนวน

เมื่อ $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i^2)$ .... (3-9)

(4) พิสัย (range) คือค่าความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของชุดข้อมูล

(5) สัมประสิทธิ์ของความแปรปรวน (coefficient of variation) การที่ข้อมูล 2 ชุดมีค่าความเบี่ยงเบนมาตรฐานเท่ากันนั้น ไม่สามารถที่จะสรุปว่า ข้อมูล 2 ชุดนั้นมีความกระจายเท่ากัน หากข้อมูลทั้ง 2 ชุดดังกล่าวมีค่าเฉลี่ยต่างกัน สิ่งหนึ่งที่สามารถวัดการกระจายเปรียบเทียบกันได้ก็คือ สัมประสิทธิ์ของความแปรปรวน แทนด้วยสัญญลักษณ์ C_V หาค่าได้จากค่าความเบี่ยงเบนมาตรฐานหารด้วยค่าเฉลี่ยเลขคณิตของข้อมูลเดียวกัน

$$C_V = \frac{\sigma}{\mu} \simeq \frac{s}{\bar{x}} \qquad \text{.......... (3-10)}$$

ค. การวัดความไม่สมมาตร (Measure of Symmetry)

การที่เส้นโค้งของการแจกแจงความน่าจะเป็น (probability distribution curve) มีความสมมาตร เบ้ซ้าย หรือ เบ้ขวา นั้นสามารถจะวัดได้ด้วยค่าความเบ้ (skewness) เมื่อแทนค่าความเบ้ของประชากรและตัวอย่างด้วยสัญญลักษณ์ $\hat{\gamma}$ และ γ ตามลำดับ สามารถหาค่าได้จากสมการคือ

$$\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^3 \qquad \text{.......... (3-11)}$$

$$\gamma = \frac{n}{(n-1)\cdot(n-2)}\sum_{i=1}^{n}(x_i-\bar{x})^3 \qquad \text{....... (3-12)}$$

$$\gamma = \frac{n^2}{(n-1)(n-2)}\left[\overline{x^3}-3\overline{x^2}\bar{x}+2\bar{x}^3\right] \qquad \text{........ (3-13)}$$

$$\overline{x^3} = \frac{1}{n}\sum_{i=1}^{n}x_i^3 \qquad \text{......... (3-14)}$$

โดยทั่วไปวิธีที่นิยมใช้สำหรับวัดความเบ้ของเส้นโค้งของการแจกแจงความน่าจะเป็นวิธีหนึ่งคือ สัมประสิทธิ์ของความเบ้ (coefficient of skew) แทนด้วยสัญญลักษณ์ $\hat{\gamma}_1$ สามารถหาค่าได้จากสมการ

$$\hat{\gamma}_1 = \frac{\hat{\gamma}}{\sigma^3} \simeq \frac{\gamma}{s^3} \qquad \text{..........} (3\text{-}15)$$

เมื่อค่า $\gamma_1 = 0$ แสดงว่าเส้นโค้งของการแจกแจงความน่าจะเป็นสมมาตร ซึ่งจะทำให้ค่าเฉลี่ย (mean) มัธยฐาน (median) และฐานนิยม (mode) มีค่าเท่ากัน ถ้าค่า $\hat{\gamma}_1 < 0$ แสดงว่าเบ้ซ้าย (left skewness) และถ้า $\hat{\gamma}_1 > 0$ แสดงว่าเบ้ขวา (right skewness)

3.1.3 การประมาณค่าพารามิเตอร์ (Estimate of Parameter)

เราสามารถจะทราบถึงคุณสมบัติเฉพาะของการแจกแจงของสถิติได้ เมื่อทราบค่าพารามิเตอร์ต่าง ๆ ของการแจกแจงของสถิตินั้น ๆ ซึ่งมีวิธีการที่จะประมาณค่าได้หลายวิธี ส่วนวิธีที่นิยมและยอมรับกันอย่างแพร่หลายคือ

1. วิธีโมเมนต์ (method of moment)
2. วิธี maximum likelihood (method of maximum likelihood)
3. วิธีกำลังสองน้อยที่สุด (method of least square)
4. วิธีกราฟ (method of graphical)

ในการวิจัยครั้งนี้จะใช้วิธีโมเมนต์เพียงวิธีเดียว ซึ่งเป็นวิธีที่ได้รับความนิยมและให้ความสะดวกในการคำนวณด้วยเครื่องไมโครคอมพิวเตอร์ (microcomputers) และให้ผลวิเคราะห์ที่ดีสำหรับข้อมูลฝน

โมเมนต์ของสถิติ (statistical moments)

เมื่อให้ $x_1, x_2, \ldots\ldots, x_n$ เป็นค่าของตัวแปรสุ่มแล้วสามารถจะหาค่าโมเมนต์อันดับที่ r ของสถิติซึ่งแบ่งออกเป็น 2 ชนิดได้

(1) โมเมนต์รอบจุดกำเนิด (origin) คือจุดที่ x = 0 สามารถจะหาค่าได้จากสมการ

$$\mu'_r = \int_{-\infty}^{+\infty} x^r p(x)\,dx \qquad \text{.........} (3\text{-}16)$$

สำหรับตัวแปรสุ่มชนิดต่อเนื่อง (continuous random variable)

$$\mu'_r = \sum_{i=1}^{n} x_i^r p(x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i^r \qquad \text{.........} (3\text{-}17)$$

สำหรับตัวแปรสุ่มชนิดไม่ต่อเนื่อง (discrete random variable)

เมื่อ $p(x)$ = ความถี่ (frequency) หรือความน่าจะเป็น (probability) ของ x แต่ละค่า

(2) โมเมนต์รอบค่าเฉลี่ย สามารถหาค่าได้จากสมการ

$$\mu_r = \int_{-\infty}^{+\infty} (x-\mu)^r p(x)\,dx \qquad \text{.........} (3\text{-}18)$$

สำหรับตัวแปรสุ่มชนิดต่อเนื่อง

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n} (x_i-\mu) \qquad \text{.........} (3\text{-}19)$$

สำหรับตัวแปรสุ่มชนิดไม่ต่อเนื่อง

สำหรับโมเมนต์อันดับที่ 1, 2 และ 3 นั้นจะมีค่าสัมพันธ์กับค่าพารามิเตอร์ต่าง ๆ ดังต่อไปนี้

$$\mu'_1 = \mu \qquad \text{.........} (3\text{-}20)$$

$$\mu'_2 = \mu_2 + {\mu'_1}^2 \qquad \text{.........} (3\text{-}21)$$

$$\mu'_3 = \mu_3 + 3\mu_2\mu'_1 + {\mu'_1}^3 \qquad \text{.........} (3\text{-}22)$$

$$\mu_1 = 0 \qquad \text{..........} (3\text{-}23)$$

$$\mu_2 = \mu_2' - \mu_1'^2 = \sigma^2 \qquad \text{..........} (3\text{-}24)$$

$$\mu_3 = \mu_3' + 3\mu_2'\mu_1' + \mu_1'^3 = \gamma\sigma^3 \qquad \text{..........} (3\text{-}25)$$

3.2 <u>แนวความคิดเกี่ยวกับความน่าจะเป็นในทางอุทก</u>

ในปัจจุบันถึงแม้ว่าสามารถจะอธิบายถึงองค์ประกอบต่าง ๆ ที่เป็นปัจจัยสำคัญบางอย่างที่ทำให้เกิดปรากฏการณ์ทางอุทกได้ แต่ก็ยังไม่สามารถทราบและอธิบายถึงความผันแปรขององค์ประกอบเหล่านั้นได้ เนื่องจากเป็นปรากฏการณ์ทางธรรมชาติซึ่งมนุษย์ยังไม่สามารถจะกำหนด ควบคุม หรือชี้ชัดถึงสาเหตุของการเกิดได้ทั้งหมด

ฉะนั้นการศึกษาเกี่ยวกับปรากฏการณ์ทางอุทกจึงมักจะศึกษาเป็นระบบ (system) คือสนใจแต่เหตุและผลของการเกิดปรากฏการณ์ต่าง ๆ เท่านั้น จะไม่สนใจรายละเอียดในระบบนั้น ๆ ด้วยเหตุนี้การศึกษาเกี่ยวกับปรากฏการณ์ทางอุทก จึงได้อาศัยการเก็บข้อมูลที่ผ่านมาเพื่อใช้ในการคาดคะเนเหตุการณ์ในอนาคต การคาดคะเนปรากฏการณ์ทางอุทกมักจะเป็นการคาดคะเนในลักษณะของความน่าจะเกิดขึ้นของขนาดของเหตุการณ์ที่น้อยกว่าหรือเท่ากับหรือมากกว่าขนาดของเหตุการณ์ที่กำหนดในช่วงระยะเวลาหนึ่ง (สำหรับการวิเคราะห์ข้อมูลทางอุทกวิทยามักจะกำหนดเป็นปี)

เมื่อกำหนดให้ X เป็นตัวแปรสุ่ม x เป็นค่าของตัวแปรสุ่ม และ A เป็นเหตุการณ์แล้ว นิยามเกี่ยวกับความน่าจะเป็นที่สำคัญทางอุทกคือ

3.2.1 ความน่าจะเป็นของเหตุการณ์ (probability of event) ทางด้านอุทกวิทยามักจะหมายถึง การแจกแจงความน่าจะเป็นสะสม (cumulative distribution) ของตัวแปรสุ่มที่มีค่าน้อยหรือเท่ากับค่าที่กำหนดในช่วงเวลาหนึ่ง

$$P(A) = P(X \leqslant x) \qquad \text{..........} (3\text{-}26)$$

$$\text{และ} \quad P(X \leqslant x) = \sum_{x_i \leqslant x} P(x_i) \qquad \text{..........} (3\text{-}27)$$

เมื่อ เป็นตัวแปรสุ่มชนิดไม่ต่อเนื่อง

$$P(X\leqslant x) \;=\; \int_{-\infty}^{+\infty} p(x)\,dx \qquad \ldots\ldots\ldots\ (3\text{-}28)$$

เมื่อ เป็นตัวแปรสุ่มชนิดต่อเนื่อง

และ $P(A)$ = ความน่าจะเป็นของเหตุการณ์

$P(X\leqslant x)$ = ความน่าจะเป็นสะสมของตัวแปรสุ่ม X ทุก ๆ ค่าที่มีค่าน้อยกว่าหรือเท่ากับ x

$P(x)$ = ฟังชั่นของความน่าจะเป็น (probability function) หรือการแจกแจงความน่าจะเป็น (probability distribution) ของตัวแปรสุ่ม

3.2.2 ค่ารอบปี (return period) คือส่วนกลับของความน่าจะเป็นของเหตุการณ์หรือค่าการแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่มที่มีค่าเท่ากับหรือมากกว่าค่าที่กำหนดในช่วงระยะเวลาเท่ากับ 1 ปี ซึ่งจะให้ความหมายว่า ในแต่ละปีนั้นเหตุการณ์ที่มีขนาดเท่ากับหรือมากกว่าที่กำหนด จะมีโอกาสเกิดขึ้นได้เท่ากับส่วนกลับของค่ารอบปี

$$T \;=\; \frac{1}{P(A)} \;=\; \frac{1}{1-P(X\leqslant x)} \;=\; \frac{1}{P(X\geqslant x)} \qquad \ldots\ldots(3\text{-}29)$$

เมื่อ T คือค่ารอบปี (ปี)

3.2.3 ความเสี่ยง (risk) หมายถึงค่าความน่าจะเป็นของเหตุการณ์ที่มีค่าเท่ากับหรือมากกว่าค่าที่กำหนดจะเกิดขึ้นอย่างน้อยหนึ่งครั้งในระยะเวลา n ปีที่ต่อเนื่อง ซึ่งสามารถประมาณค่าได้จากสมการ

$$R \;=\; 1-(1-1/T)^{n} \qquad \ldots\ldots\ldots(3\text{-}30)$$

เมื่อ R = ความเสี่ยง

3.3 ตัวคูณค่าความถี่ (Frequency Factor)

Chow (1964) ได้เสนอหลักการว่า เมื่อ x คือค่าของเหตุการณ์ใด ๆ สามารถที่จะหาค่าได้ในรูปของการเบี่ยงเบนไปจากค่าเฉลี่ย และได้เสนอสมการในรูป

$$x = \mu + \Delta x \qquad \text{.........} (3-31)$$

โดยให้ x แทนค่าของเหตุการณ์ใด ๆ และ Δx เป็นส่วนที่เบี่ยงเบนไปจากค่าเฉลี่ย และได้ตั้งสมมุติฐานว่าค่า Δx มีค่าเท่ากับผลคูณของค่าเบี่ยงเบนมาตรฐาน σ กับค่าตัวคูณค่าความถี่ (K)

$$\Delta x = \sigma K \qquad \text{.........} (3-32)$$

ดังนั้นจะได้สมการทั่วไปสำหรับการคำนวณค่าเหตุการณ์ที่รอบปีใด ๆ เมื่อแทนค่ารอบปีและค่าเหตุการณ์ของรอบปีนั้นด้วยสัญลักษณ์ T และ x_T ตามลำดับคือ

$$x_T = \mu + K\sigma \qquad \text{.........} (3-33)$$

$$x_T = m_1' + K\sqrt{m_2} \qquad \text{.........} (3-34)$$

เมื่อให้ m_1' และ m_2 แทนค่าเฉลี่ยและค่าความแปรปรวนของตัวอย่าง ซึ่งสามารถประมาณค่าได้โดยวิธีโมเมนต์ สำหรับตัวคูณค่าความถี่ (K) นั้นสามารถที่จะพิสูจน์หาได้ตามแต่ละทฤษฎีของการแจกแจงความน่าจะเป็น

3.4 ค่าความคลาดเคลื่อนมาตรฐานของการกะประมาณ (Standard Error of Estimate)

การวัดความแปรปรวนของการคำนวณค่าเหตุการณ์ใด ๆ สามารถจะวัดได้ด้วยค่าความคลาดเคลื่อนมาตรฐาน เป็นค่าที่บอกให้ทราบว่าผลการวิเคราะห์ข้อมูลในแต่ละชุดข้อมูลมีความคลาดเคลื่อนของการกะประมาณและค่าจากการสังเกต (observes) ที่รอบปีต่าง ๆ มากน้อยเพียงใด ค่าความคลาดเคลื่อนมาตรฐานของการกะประมาณสามารถประมาณค่าได้จากสมการ

$$S_T = \left[\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \right]^{1/2} \qquad \text{..........} \quad (3\text{-}35)$$

เมื่อ $\hat{x}_i$ คือค่าเหตุการณ์ที่ความถี่ต่าง ๆ ซึ่งคำนวนได้ตามทฤษฎีการแจกแจงความน่าจะเป็น ของแต่ละทฤษฎี และ S_T แทนค่าความคลาดเคลื่อนมาตราฐานของการกะประมาณ ของทฤษฎีการแจกแจงความน่าจะเป็นแต่ละทฤษฎี

Kite (1977) แสดงวิธีการประมาณค่าความคลาดเคลื่อนมาตราฐานของการกะประมาณ ซึ่งสัมพันธ์กับค่าโมเมนต์ 3 อันดับ และค่ารอบปีคือ $S_T = f\ (m_1', m_2, m_3, T)$ เมื่อ m_1' คือค่าโมเมนต์อันดับที่ 1 รอบจุดกำเนิด, m_2 และ m_3 คือโมเมนต์อันดับที่ 2 และ 3 รอบค่าเฉลี่ยของตัวอย่าง และ Kite(1977) ยังได้พิสูจน์ให้เห็นว่า

$$S_T^2 = \frac{\mu_2}{n}\Big\{1+K\gamma_1+\frac{K^2}{4}\left[\gamma_2-1\right]+\frac{\partial K}{\partial \gamma_1}\left[2\gamma_2-3\gamma_1^2-6+K\left(\gamma_3-\frac{6}{4}\gamma_1\gamma_2-\frac{10}{4}\gamma_1\right)\right]$$

$$+\left(\frac{\partial K}{\partial \gamma_1}\right)^2\left[\gamma_4-3\gamma_3\gamma_1-6\gamma_2+\frac{9}{4}\gamma_1^2\gamma_2+\frac{35}{4}\gamma_1^2+9\right]\Big\} \qquad \text{..........} \quad (3\text{-}36)$$

และ $$\gamma_1 = \mu_3/\mu_2^{3/2} \qquad \text{..........} \quad (3\text{-}37)$$

$$\gamma_2 = \mu_4/\mu_2^{2} \qquad \text{..........} \quad (3\text{-}38)$$

$$\gamma_3 = \mu_5/\mu_2^{5/2} \qquad \text{..........} \quad (3\text{-}39)$$

$$\gamma_4 = \mu_6/\mu_2^{3} \qquad \text{..........} \quad (3\text{-}40)$$

สมการโดยทั่วไป สำหรับการหาค่าความคลาดเคลื่อนมาตราฐานของการกะประมาณ คือ

$$S_T = \delta\sqrt{\frac{\mu_2}{n}} \qquad \text{..........} \quad (3\text{-}41)$$

เมื่อ δ คือค่า พารามิเตอร์สำหรับค่าความคลาดเคลื่อนมาตราฐาน (parameter for use in standard error) ซึ่งสามารถหาค่าได้จากตารางของค่า δ ของการแจกแจกความน่าจะเป็นแต่ละชนิด และสำหรับการแจกแจงความน่าจะเป็นที่มีค่า K ไม่ขึ้นกับสัมประสิทธิ์ของความเบ้แล้ว สามารถหาค่าของ δ ได้จากสมการ

$$\delta = \{1+K\gamma_1+\frac{K^2}{4}\left[\gamma_2-1\right]\}^{1/2} \qquad \text{............ (3-42)}$$

3.5 <u>ทฤษฎีการแจกแจงความน่าจะเป็น</u>

ทฤษฎีการแจกแจงความน่าจะเป็นที่ถูกนำมาใช้ สำหรับการวิเคราะห์ความถี่ของการเกิดฝนสูงสุดนั้น มีอยู่เป็นจำนวนมากแต่ที่นิยมใช้กันอย่างแพร่หลายนั้นมีอยู่ไม่มากนัก ดังจะกล่าวถึงต่อไป

3.5.1 การแจกแจงแบบทรันเคทนอร์มอล (Truncated Normal Distribution)

การแจกแจงแบบนี้มีฟังชั่นของความน่าจะเป็นคือ

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad \text{............. (3-43)}$$

เมื่อ x มีค่าตั้งแต่ $-\infty$ ถึง $+\infty$, μ และ σ คือค่าเฉลี่ยและความเบี่ยงเบนมาตราฐานของประชากร และสามารถประมาณค่าได้จากสมการ

$$\mu = \mu_1' \qquad \text{.............. (3-44)}$$

$$\sigma = \sqrt{\mu_2} \qquad \text{.............. (3-45)}$$

และเมื่อใช้การประมาณค่าของเหตุการณ์ด้วยสมการโดยทั่วไป ของการแจกแจงความน่าจะเป็นตามวิธีการของ Chow (1964) ; $x_T = \mu+K\sigma$ ค่า K จะมีค่าเท่ากับค่าความเบี่ยงเบนมาตราฐานปกติ (standard normal deviation) ซึ่งแทนด้วยสัญญาลักษณ์ t' และสามารถประมาณค่าโดยวิธี โพลิโนเมียล (polynomial approximation) ได้จากสมการ

$$K = t' \approx w-\frac{c_0+c_1w+c_2w^2}{1+d_1w+d_2w^2+d_3w^3} \qquad \text{.............. (3-46)}$$

$c_0 = 2.515517$; $c_1 = 0.802853$; $c_2 = 0.010328$

$d_1 = 1.432788$; $d_2 = 0.189269$; $d_3 = 0.001308$

$w = \sqrt{\ln\{1/P(t')^2\}} \qquad 0 < P(t') \le 0.5$

เมื่อ $P(t') > 0.5$ แล้ว $P(t')$ จะมีค่าเท่ากับ $1-P(t')$

ฉนั้นเมื่อค่ารอบปี T มีค่าตั้งแต่ 2 ปีขึ้นไป $w = \sqrt{\ln(T)}$

ค่าความคลาดเคลื่อน มาตราฐานของการกะประมาณ สามารถหาค่าได้จากสมการ

$$S_T = \delta\left[\sigma/\sqrt{n}\right] \quad \text{.......... (3-47)}$$

$$\delta = \left[1+\gamma_1 K+(\gamma_2-1)\frac{K^2}{4}\right]^{1/2} \quad \text{.......... (3-48)}$$

สำหรับการแจกแจงแบบทรันเคทนอร์มอล โมเมนต์รอบค่าเฉลี่ยอันดับต่าง ๆ มีค่าดังต่อไปนี้ $\mu_2 = \sigma^2$; $\mu_3 = 0$; $\mu_4 = 3\sigma^4$ และ $K = t'$ และจากความสัมพันธ์ ในสมการ (3-37) และ (3-38) สามารถจัดรูปในสมการ (3-48) ได้ใหม่ให้ง่ายขึ้นคือ

$$\delta = \left[1+(t'^2/2)\right]^{1/2} \quad \text{............ (3-49)}$$

3.5.2 การแจกแจงแบบลอกนอร์มอลชนิด 2 พารามิเตอร์ (2-Parameter Lognormal Distribution)

เป็นการแจกแจงที่ดัดแปลงมาจาก การแจกแจงแบบทรันเคทนอร์มอล โดยแปลงข้อมูลให้อยู่ในรูปของค่าลอกคือ $y = \ln(x)$ และมีฟังชั่นของความน่าจะเป็นดังสมการ

$$p(x) = \frac{1}{x\sigma_y\sqrt{2\pi}}\exp\{-\frac{\left[\ln(x)-\mu_y\right]^2}{2\sigma_y^2}\} \quad \text{....... (3-50)}$$

เมื่อ μ_y และ σ_y คือค่าเฉลี่ยและค่าความเบี่ยงเบนมาตราฐานของ y

โมเมนต์อันดับต่าง ๆ ของการแจกแจงแบบลอกนอร์มอลชนิด 2 พารามิเตอร์คือ

$$\mu'_r = \int_0^\infty x^r \frac{1}{x\sigma_y\sqrt{2\pi}}\exp\{-\frac{\left[\ln(x)-\mu_y\right]^2}{2\sigma_y^2}\}dx \quad \text{........ (3-51)}$$

เมื่อกำหนดให้ $\omega = \left[\ln(x)-r\sigma_y^2-\mu_y\right]/\sigma_y$ แล้วแทนค่าลงในสมการ (3-51) จะได้

$$\mu'_r = \exp\left[r\mu_y+\frac{r^2\sigma_y}{2}\right]\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{\omega^2}{2}\right]d\omega \quad \text{.......... (3-52)}$$

จากสมการ (3-52) เมื่อจำนวนหลังมีค่าเท่ากับ 1 ฉะนั้น

$$\mu_r' = \exp\left[r\mu_y + \frac{r^2\sigma_y^2}{2}\right] \quad \text{..........(3-53)}$$

$$\mu_1' = \mu = \exp\left[\mu_y + \frac{\sigma_y^2}{2}\right] \quad \text{..........(3-54)}$$

$$\mu_2 = \sigma^2 = \left[\exp(\sigma_y^2)-1\right].\mu_1'^2 \quad \text{..........(3-55)}$$

$$\mu_3 = \left[\exp(3\sigma_y^2)-3\exp(\sigma_y^2)+2\right].\exp(3\mu_y + \frac{3\sigma_y^2}{2}) \quad \text{...(3-56)}$$

จากสมการ (3-54), (3-55) และ (3-56) สามารถหาค่าสัมประสิทธิ์ของความเบี่ยงเบน และสัมประสิทธิ์ของความเบ้ได้จากสมการ

$$Z = \left[\exp(\sigma_y^2)-1\right]^{1/2} \quad \text{...........(3-57)}$$

$$\gamma_1 = \frac{\exp(3\sigma_y^2)-3\exp(\sigma_y^2)+2}{\left[\exp(\sigma_y^2)-1\right]^{3/2}} \quad \text{...........(3-58)}$$

แทนค่า Z ลงในสมการ (3-58) จะได้ $\gamma_1 = 3Z+Z^3$

การวิเคราะห์ความถี่โดยสมการทั่วไปของ ChOw (1964) ได้โดยอาศัยหลักการทำนองเดียวกับการแจกแจงแบบทรันเคทนอร์มอล และแปลงค่าข้อมูลให้เป็นค่าลอกคือ

$$\ln(x_T) = y_T = \mu_y + t'\sigma_y \quad \text{.............(3-59)}$$

แทนค่าจากสมการ (3-54) และ (3-55) ลงในสมการ (3-59) จะได้ว่า

$$\exp(y_T) = \exp(\mu_y + \frac{\sigma_y^2}{2})\{1+K\left[\exp(\sigma_y^2)-1\right]^{1/2}\} \quad \text{........(3-60)}$$

$$K = \frac{\exp(y_T-\mu_y-\frac{\sigma_y^2}{2})-1}{\left[\exp(\sigma_y^2)-1\right]^{1/2}} \quad \text{.............(3-61)}$$

จากสมการ (3-59) $y_T-\mu_y = t'\sigma_y$ จะได้ว่า

$$K = \frac{\exp(\sigma_y t' - \frac{\sigma_y^2}{2}) - 1}{\left[\exp(\sigma_y^2) - 1\right]^{1/2}} \qquad \text{.......... (3-62)}$$

จากสมการ (3-57) และ (3-62) จะได้ว่า

$$K = \frac{\exp\{\left[\ln(1+Z^2)\right]^{1/2} t' - \left[\ln(1+Z^2)\right]/2\} - 1}{Z} \qquad \text{.......(3-63)}$$

ความคลาดเคลื่อนมาตราฐานของการกะประมาณ สำหรับการแจกแจงแบบลอกนอร์มอล ชนิด 2 พารามิเตอร์ สามารถหาค่าในหน่วยลอก (Logarithmic units) ได้จากสมการ

$$S_T = \delta\sigma_y/\sqrt{n} \qquad \text{..............(3-64)}$$

$$\delta = \{1+\gamma_1 K + \left[\gamma_2 - 1\right]\frac{K^2}{4}\}^{1/2} \qquad \text{..............(3-65)}$$

เมื่อ σ_y, γ_1 และ γ_2 เป็นค่าในหน่วยลอก และ $K=t'$ สมการ (3-65) สามารถจัดรูปให้ง่ายขึ้นได้เป็น

$$\delta = (1+t'^2/2)^{1/2} \qquad \text{...............(3-66)}$$

สำหรับความคลาดเคลื่อนมาตราฐานของการกะประมาณ ในหน่วยเชิงเส้น (linear units) สามารถหาได้จากสมการ

$$s_T = \delta_y \sigma/\sqrt{n} \qquad \text{...............(3-67)}$$

เมื่อ σ เป็นค่าในหน่วยเชิงเส้น และสำหรับ γ_2 ในหน่วยลอกนั้น สามารถหาค่าได้จากสมการ

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} = \left[\exp(\sigma_y^2/2)\right]^8 + 2\left[\exp(\sigma_y^2/2)\right]^6 + 3\left[\exp(\sigma_y^2/2)\right]^4 - 3 \quad \text{....(3-68)}$$

จากสมการ (3-58), (3-65) และ (3-68) จะได้ว่า

$$\delta_y = \left[1+(Z^3+3Z)K+(Z^8+6Z^6+15Z^4+16Z^2+2)K^2/4\right]^{1/2} \quad \text{.......(3-69)}$$

3.5.3 การแจกแจงแบบเพียร์สันชนิดที่ 3

การแจกแจงแบบเพียร์สันชนิดที่ 3 นี้มีฟังชั่นของความน่าจะเป็น คือ

$$p(x) = \frac{1}{\alpha\Gamma(\beta)}\left[\frac{x-\gamma}{\alpha}\right]^{\beta-1}\exp\left[-\frac{x-\gamma}{\alpha}\right] \quad ......(3-70)$$

เมื่อ $\gamma,\beta,$ และ α คือ shape, scalc และ location พารามิเตอร์ตามลำดับและ $\Gamma(\beta)$ คือ แกมม่าฟังชั่น (gamma function)

โดยอาศัยวิธีการของโมเมนต์ Bobee และ Robitaille ได้แสดงให้เห็นว่าสามารถประมาณค่าพารามิเตอร์ต่าง ๆ สำหรับการแจกแจงแบบเพียร์สันชนิดที่ 3 ได้ดังสมการ [Kite (1977)]

$$\alpha = \sigma/\sqrt{\beta} \quad ....(3-71)$$

$$\gamma = \mu-\sigma\sqrt{\beta} \quad ....(3-72)$$

$$\beta = (2/\gamma_1)^2 \quad ....(3-73)$$

$$\gamma_1 = \hat{\gamma}_1\frac{\sqrt{n(n-1)}}{n-2}\left[1+\frac{8.5}{n}\right] \quad ....(3-74)$$

ค่า K สำหรับสมการโดยทั่วไปของการวิเคราะห์ความถี่ของการแจกแจงแบบนี้สามารถประมาณค่าได้จากสมการ

$$K \simeq t'+(t'^2-1)\frac{\gamma_1}{6}+\frac{1}{3}(t'^3-6t')\left(\frac{\gamma_1}{6}\right)^2-(t'^2-1)\left(\frac{\gamma_1}{6}\right)^3+t'\left(\frac{\gamma_1}{6}\right)^4+\frac{1}{3}\left(\frac{\gamma_1}{6}\right)^5 \quad ...(3-75)$$

และค่าความคลาดเคลื่อนมาตราฐานของการกะประมาณ สามารถหาค่าได้จากสมการ (3-36)-(3-39) โดยจัดให้ γ_2 และ γ_3 อยู่ในรูปของ γ_1 คือ

$$\gamma_2 = 3(1+\gamma_1^2/2) \text{ และ } \gamma_3 = \gamma_1(10+3\gamma_1^2) \quad ...(3-76)$$

$$S_T^2 = \frac{\mu}{n}2\left\{1+K\gamma_1+\frac{K^2}{2}\left[\frac{3\gamma_1^2}{4}+1\right]+3K\frac{\partial K}{\partial\gamma_1}\left[\gamma_1+\frac{\gamma_1^3}{4}\right]+3\left(\frac{\partial K}{\partial\gamma_1}\right)^2\left[2+3\gamma_1^2+\frac{5\gamma_1^4}{8}\right]\right\} \quad ...(3-77)$$

และ $\frac{\partial K}{\partial \gamma_1} \approx \frac{t'^2-1}{6}+\frac{4(t'^3-6t)}{6^3}\gamma_1-\frac{3(t'^2-1)}{6^3}\gamma_1^2+\frac{4t'}{6^4}\gamma_1^3-\frac{10}{6^6}\gamma_1^4$ .... (3-78)

3.5.4 การแจกแจงแบบลอกเพียร์สันชนิดที่ 3 (Log-Pearson Type III Distribution)

การแจกแจงแบบนี้อาศัยหลักการของการแจกแจงแบบเพียร์สันชนิดที่ 3 และเปลี่ยนข้อมูลให้เป็นค่าลอก คือ

$$y_T = \ln(x_T) = \mu_y + K\sigma_y \quad \text{...... (3-79)}$$

และประมาณค่าความคลาดเคลื่อนมาตราฐาน ของการกะประมาณสามารถประมาณค่าในหน่วยเชิงเส้นได้จากสมการ

$$S_{T,x} = \{x_T[\exp(S_{T,y})-\exp(-S_{T,y})]\}/2 \quad \text{......(3-80)}$$

เมื่อ $S_{T,x}$ และ $S_{T,y}$ คือค่าความคลาดเคลื่อนมาตราฐานของการกะประมาณในหน่วยเชิงเส้นและหน่วยลอกตามลำดับ

3.5.5 การแจกแจงแบบกัมเบล (Gumbel or Type I Extremal Distribution)

มีฟังชั่งของความน่าจะเป็นคือ

$$p(x) = \alpha\cdot\exp\{-\alpha(x-\beta) - \exp[-\alpha(x-\beta)]\} \quad \text{......(3-81)}$$

เมื่อ α และ β คือค่า scale และ location (central value) พารามิเตอร์ตามลำดับ ซึ่งสามารถหาค่าได้จากสูตร

$$\alpha \approx 1.2825/\sigma \quad \text{........ (3-82)}$$

$$\beta = \mu-0.450\sigma \quad \text{........ (3-83)}$$

ค่า K สำหรับการแจกแจงแบบกัมเบลนี้ สามารถหาค่าได้จากสมการ

$$K = \frac{Y_T-\bar{Y}_n}{S_n} \quad \text{......... (3-84)}$$

$$Y_T = -\ln\{-\ln[(T-1)/T]\} \quad \text{.......} \; (3\text{-}85)$$

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_m \quad \text{.......} \; (3\text{-}86)$$

$$Y_m = -\ln\{-\ln[(n+1-m)/(n+1)]\} \quad \text{.....} \; (3\text{-}87)$$

$$S_n = [\frac{1}{n}\sum_{i=1}^{n}(Y_m-\bar{Y}_n)^2]^{1/2} \quad \text{.......} \; (3\text{-}88)$$

เมื่อ T = ค่ารอบปีตามต้องการ (ปี)

n = จำนวนของข้อมูล

m = ลำดับที่ของค่าเหตุการณ์ของข้อมูลจากมากไปน้อย

สำหรับค่าความคลาดเคลื่อนมาตรฐานของการกะประมาณ หาได้จากสมการ (3-41) และการแจกแจงแบบนี้กำหนดให้ $\gamma_1 = 1.1396$ และ $\gamma_2 = 5.4002$ ดังนั้น

$$S_T^2 = \frac{\sigma^2}{n}[1+1.1396K+1.1000K^2] \quad \text{.......} \; (3\text{-}89)$$

3.6 <u>การทดสอบความเหมาะสม</u> (Test of Goodness of Fit) ของการแจกแจงแบบต่าง ๆ

การทดสอบความเหมาะสมของการแจกแจงความน่าจะเป็นแบบต่าง ๆ เพื่อหาทฤษฎีที่เหมาะสมกับกลุ่มตัวอย่างของเหตุการณ์ทางอุทกวิทยานั้น เป็นสิ่งสำคัญมากอย่างหนึ่งที่ต้องกระทำก่อนที่จะตัดสินใจเลือกใช้ทฤษฎีการแจกแจงที่เหมาะสมที่สุดในการวิเคราะห์ เพื่อคาดคะเนความน่าจะเป็นของเหตุการณ์ขนาดต่าง ๆ ตามวิธีการของทฤษฎีการแจกแจงความน่าจะเป็น ความผิดพลาดหรือความไม่เหมาะสมในการคาดคะเนอาจจะเกิดขึ้นได้ หากเลือกทฤษฎีการแจกแจงความน่าจะเป็นที่ไม่เหมาะสมกับการแจกแจงของค่าเหตุการณ์ทางอุทกวิทยานั้น ๆ เพราะทฤษฎีการแจกแจงความน่าจะเป็นที่มีอยู่ในปัจจุบันนั้น มีอยู่เป็นจำนวนมากที่ได้ถูกประยุกต์มาใช้สำหรับการคาดคะเนค่าเหตุการณ์ ของปรากฏการณ์ทางอุทกวิทยา ซึ่งแต่ละทฤษฎีนั้นอาจจะมีความเหมาะสมเฉพาะพื้นที่, ประเภท และชุดของข้อมูลของปรากฏการณ์ทางอุทกนั้น ๆ Kite (1977) ได้กล่าวถึงผลการวิจัยเกี่ยวกับการทดสอบความเหมาะสมของการแจกแจงความน่าจะเป็นของบุคคลอื่น ๆ เพื่อแสดงให้เห็นว่าจะไม่มีทฤษฎีการแจกแจงความน่าจะเป็นแบบใดแบบหนึ่งที่จะมีความเหมาะสมกับปรากฏการณ์ทางอุทกได้ทั้งหมด ในทุกพื้นที่, ทุกประเภท และทุกชุดข้อมูล ฉะนั้นจึงต้องมีการทดสอบความเหมาะสมของทฤษฎีการแจกแจงความน่าจะเป็น ก่อนเลือกมาใช้ในการวิเคราะห์

สำหรับวิธีการทดสอบความเหมาะสมของทฤษฎีการแจกแจงความน่าจะเป็น ที่นิยมใช้กันอย่างแพร่หลายมากในปัจจุบันคือ Chi-Square test, Kolmogorov-Smirnov test, และ Sum of square of differences between calculated and observed event ซึ่งมีรายละเอียดดังจะได้กล่าวถึงโดยสังเขปต่อไป

3.6.1 Chi-Square test

การทดสอบความเหมาะสมของการแจกแจงวรรคความถี่ ที่ได้จากการสังเกต (observed number of events) และความถี่ที่คาดว่าจะได้ (exspected number of events) จะมีค่าความเหมาะสมเพียงใดนั้นขึ้นอยู่กับ

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \qquad \text{........ (3-9o)}$$

เมื่อ χ^2 เป็นค่าของตัวแปรสุ่ม χ^2 ซึ่งมีการแจกแจงใกล้เคียงกับการแจกแจงไคสแควมากที่สุด สัญญลักษณ์ O_j และ E_j คือค่าความถี่ที่ได้จากการสังเกตและความถี่ที่คาดว่าจะได้ตามทฤษฎีตามลำดับ และ k คือ จำนวนช่างชั้น (number of class intervals)

การกำหนดขนาดและระยะของค่าช่วงชั้นนั้น ยังไม่มีทฤษฎียืนยันสำหรับการทดสอบแบบไคสแคว Markovic (1965) ได้กล่าวว่า นักสถิติส่วนมากได้แนะนำว่าจำนวนช่วงชั้นไม่ควรเกินกว่า 20 ช่วงชั้น และไม่ควรน้อยกว่า 10 ช่วงชั้น เพราะถ้ามีจำนวนช่วงชั้นมากเกินไปจะทำให้ในบางช่วงชั้นไม่มีค่าของเหตุการณ์อยู่เลย ซึ่งจะทำให้ค่าที่ได้ผิดปกติไป และถ้าหากจำนวนช่วงชั้นมีน้อยเกินไป จะทำให้ผลการทดสอบไม่เด่นชัด แต่ในทางปฏิบัติโดยทั่วไปนั้น นักสถิติมักกำหนดให้มีค่าที่คาดว่าจะได้ในแต่ละช่วงชั้น ไม่น้อยกว่า 5 Yevjevich (1972) ได้เสนอว่าการกำหนดจำนวนช่วงชั้นนั้นโดยทั่วไป สำหรับการทดสอบความเหมาะสมในทางอุทกวิทยานั้น ควรให้มีจำนวนช่วงชั้นไม่น้อยกว่า 5 และค่าที่คาดว่าจะได้ในแต่ละช่วงชั้นไม่ควรน้อยกว่า 5 ด้วยเช่นกัน Haan (1977) เสนอว่าค่าที่คาดว่าจะได้ในแต่ละช่วงชั้น นักสถิติส่วนมากมักจะกำหนดให้มีค่า ไม่น้อยกว่า 3 หรือ 5

3.6.1.1 การแบ่งช่วงชั้น มีวิธีการแบ่งช่วงชั้นในการทดสอบอยู่ 2 วิธีคือ

ก. กำหนดให้มีการเพิ่มค่าของเหตุการณ์ในแต่ละช่วงชั้นเท่า ๆ กัน (equal lenghts)

ข. กำหนดให้มีการเพิ่มค่าความน่าจะเป็น ในแต่ละช่วงชั้นเท่าๆกัน (equal probabilities) ซึ่งจะทำให้ค่าที่คาดว่าจะได้ในแต่ละช่วงชั้นมีจำนวนเท่ากัน คือ

$$E_j = \frac{n}{k} \qquad \text{.......(3-91)}$$

เมื่อ n คือจำนวนของเหตุการณ์ (sample size)

สำหรับการวิจัยในครั้งนี้จะเลือกใช้วิธีการทดสอบ โดยแบ่งช่วงชั้นให้มีการเพิ่มของความน่าจะเป็นในแต่ละช่วงชั้นเท่า ๆ กัน

3.6.1.2 การหาค่าลิมิตของช่วงชั้น (class limits) ทฤษฎีการแจกแจงที่เลือกมาใช้ในการวิจัยครั้งนี้ สามารถหาลิมิตของช่วงชั้นในการทดสอบแบบไคสแคว แต่ละทฤษฎีได้ดังต่อไปนี้

ก. การแจกแจงแบบนอร์มอล

$$CL = \bar{x}+t'S \qquad \text{......(3-92)}$$

เมื่อ CL คือค่าลิมิตของช่วงชั้น , $\bar{x}$ และ S คือค่าเฉลี่ย และค่าเบี่ยงเบนมาตราฐานของกลุ่มตัวอย่างของค่าเหตุการณ์, และ t' คือค่าความเบี่ยงเบนมาตราฐานปกติ ที่ความน่าจะเป็น P ตามกำหนดของแต่ละช่วงชั้น

ข. การแจกแจงแบบลอกนอร์มอลชนิด 2-พารามิเตอร์

$$CL = \exp(\bar{y}+t'S_y) \qquad \text{.......(3-93)}$$

เมื่อ $\bar{y}$ และ S_y คือค่าเฉลี่ยและค่าเบี่ยงเบนมาตราฐานของค่าเหตุการณ์ในหน่วยลอก $[y=\ln(x)]$

ค. การแจกแจงแบบเพียร์สันชนิดที่ 3

$$CL = \bar{x}+\left[\frac{\chi^2\gamma_1}{4}-\frac{2}{\gamma_1}\right]S \qquad \text{....... (3-94)}$$

เมื่อ χ^2 คือค่าของไคสแคว ที่ความน่าจะเป็น P ของแต่ละช่วงชั้นและค่า degrees of freedom = $8/\gamma_1^2$ เมื่อ γ_1 คือสัมประสิทธิ์ของความเบ้ของกลุ่มตัวอย่าง

ง. การแจกแจงแบบลอกเพียร์สันชนิดที่ 3

$$CL = \exp\{\bar{y}+[\frac{\chi^2\gamma_y}{4}y-\frac{2}{\gamma_y}]S_y\} \quad \ldots\ldots\ldots (3\text{-}95)$$

เมื่อ γ_y คือค่าสัมประสิทธิ์ของความเบ้ของกลุ่มตัวอย่างในหน่วยลอก

จ. การแจกแจงแบบกัมเบล

$$CL = \bar{x} + [\frac{y_m-\bar{y}_n}{S_n}]S \quad \ldots\ldots\ldots (3\text{-}96)$$

เมื่อ $y_m = -\ln[-\ln(P)]$

การทดสอบความเหมาะสมของการแจกแจงแบบไคสแควร์ มีเกณฑ์พิจารณาความเหมาะสม คือค่า χ^2 จะมีค่าน้อย เมื่อความถี่ที่ได้จากการสังเกตและความถี่ที่คาดว่าจะได้ตามทฤษฎี มีค่าใกล้เคียงกัน และ χ^2 จะมากเมื่อความถี่ต่างกันมาก ดังนั้นขอบเขตของความแตกต่างที่ยอมรับได้จะต้องให้ค่า χ^2 น้อยกว่าค่า χ^2_α ซึ่งเป็นค่าที่ได้จากตารางการแจกแจงไคสแควร์ เมื่อระดับความมีนัยสำคัญ (level of significance) เท่ากับ α และองศา ของความเสรี (degree of freedom); $\nu = k-p-1$ เมื่อ p คือ จำนวนของพารามิเตอร์ที่ใช้ในการคำนวนของการแจกแจงแต่ละชนิด และการตัดสินด้วยเกณฑ์ขอบเขตที่ยอมรับได้นี้จะใช้ได้ก็ต่อเมื่อความถี่ที่คาดว่าจะได้ในแต่ละช่วงชั้นมีค่าไม่น้อยกว่า 5 [จุฬาลงกรณ์มหาวิทยาลัย คณาจารย์ ภาควิชาคณิตศาสตร์ (2523), ความน่าจะเป็นและสถิติ] และการทดสอบแบบไคสแควร์นี้ ยังใช้ในการทดสอบความเหมาะสมเปรียบเทียบของข้อมูลแต่ละชุดได้อีกด้วย เมื่อผลการทดสอบสามารถยอมรับความเหมาะสมของทฤษฎีการแจกแจงมากกว่า 1 ทฤษฎี ของข้อมูลชุดนั้น [Kite (1977)] และใช้เกณฑ์ตัดสินว่าทฤษฎีที่เหมาะสมที่สุดจะมีค่า χ^2 น้อยที่สุด

3.6.2 Kolmogorov - Smirnov test

การทดสอบแบบนี้ เป็นการทดสอบที่ใช้เกณฑ์ผลต่างสูงสุด โดยไม่คิดเครื่องหมายของความน่าจะเป็นสะสมที่แตกต่างกัน ของค่าที่คาดว่าจะได้และค่าจากการสังเกต ในแต่ละช่วงชั้น ซึ่งสามารถหาค่าได้จากสมการ

$$D_n = \max[P(x)-P_o(x)] \quad \ldots\ldots\ldots (3\text{-}97)$$

เมื่อ D_n คือค่า Kolmogorov-Smirnov statistic

$P(x)$ คือค่าความน่าจะเป็นสะสม ของค่าจากการสังเกตในแต่ละช่วงชั้น ซึ่งสามารถหาได้จากสมการ

$$P(x) = \frac{n_j}{n} \qquad \text{.........(3-98)}$$

เมื่อ n_j คือค่าความถี่สะสมของค่าจากการสังเกต ในช่วงชั้นที่ j

$P_o(x)$ คือค่าความน่าจะเป็นสะสมที่คาดว่าจะได้ในแต่ละช่วงชั้น ซึ่งสามารถหาค่าได้จากสมการ

$$P_o(x) = \frac{1}{k}, \frac{2}{k}, \ldots \qquad \text{.........(3-99)}$$

เมื่อ k คือ จำนวนช่วงชั้น สำหรับการทดสอบความเหมาะสมและไม่มีข้อกำหนดเกี่ยวกับขนาดและระยะของช่วงชั้น ในการทดสอบด้วยวิธีนี้ สามารถที่จะกำหนดได้โดยอิสระ ซึ่งบางครั้งเรียกว่าเป็น "การทดสอบแบบอิสระ" (free test)

การหาลิมิตของช่วงชั้นสามารถคำนวนได้เช่นเดียวกับวิธีการทดสอบแบบไคสแคว

และการทดสอบแบบ Kolmogorov-Smirnov นี้สามารถกระทำได้อีกวิธีหนึ่ง โดยการหาค่าความแตกต่างสูงสุด ของความน่าจะเป็นของค่าจากการสังเกตกับค่าที่คำนวนได้จากทฤษฎี ซึ่งมีขนาดของเหตุการณ์เท่ากัน

การทดสอบแบบนี้ สามารถตัดสินความเหมาะสมได้ในทำนองเดียวกับการทดสอบแบบไคสแคว คือกำหนดขอบเขตที่ยอมรับได้ของ D_n ว่าจะต้องไม่เกิน D_α ซึ่งเป็นค่าที่หาได้จากตารางค่าวิกฤติแบบ Kolmogorov-Smirnov ที่มีค่าขึ้นอยู่กับระดับความมีนัยสำคัญ (α) และขนาดของกลุ่มตัวอย่างที่ทำการทดสอบความเหมาะสม และหากผลการทดสอบมีการยอมรับทฤษฎีการแจกแจงมากกว่า 1 ทฤษฎี สำหรับข้อมูลแต่ละชุดที่ระดับความมีนัยสำคัญที่กำหนดแล้ว สามารถจะทำการทดสอบความเหมาะสมเปรียบเทียบได้ โดยหลักการที่ว่า การแจกแจงที่เหมาะสมที่สุดจะมีค่า D_n น้อยที่สุด [Kite(1977)]

3.6.3 Sum of square of differences between calculated an observed events

การทดสอบแบบนี้เป็นทางเลือกอีกทางหนึ่งของการทดสอบความเหมาะสม นอกเหนือจาก

การทดสอบแบบ Chi - squareและ Kolmogorov-Smirnov โดยอาศัยหลักการของผลรวมของผลต่างกำลังสองระหว่างขนาดของเหตุการณ์ที่ได้จากการคำนวนกับค่าจากการสังเกตทุกค่า ที่มีความน่าจะเป็นเท่ากับค่าจากการสังเกต โดยอาศัยหลักการกำลังสองน้อยที่สุด (least squares) สามารถหาค่าได้จากสมการ

$$SE_j = \left[\frac{\sum_{i=1}^{n} (x_i-\hat{y}_i)^2}{n-m_j}\right]^{1/2} \qquad \ldots\ldots\ldots(3\text{-}100)$$

เมื่อ $x_i (i = 1,2,3,\ldots,n)$ คือค่าเหตุการณ์ที่ได้จากการสังเกต ที่ได้ทำการจดบันทึกไว้มีจำนวน n ค่า และ $\hat{y}_j$ คือค่าเหตุการณ์ที่ได้จากการคำนวนทุก ๆ ค่าที่มีความน่าจะเป็นเท่ากับค่าจากการสังเกต, และ m_j คือจำนวนของพารามิเตอร์ สำหรับการคำนวนของการแจกแจงชนิดที่ j การทดสอบวิธีนี้มีข้อเสียคือผลที่ได้จากการคำนวนจะขึ้นอยู่กับวิธี plotting position ที่ใช้ในการคำนวนค่าความน่าจะเป็นของค่าเหตุการณ์ทุก ๆ ค่าของชุดข้อมูลที่มีอยู่และการทดสอบวิธีนี้จะใช้ได้เฉพาะการพิจารณาค่าความเหมาะสมเปรียบเทียบเท่านั้น โดยใช้เกณฑ์การพิจารณาว่า ทฤษฎีการแจกแจงที่เหมาะสมที่สุดจะให้ค่าความแตกต่าง (SE) น้อยที่สุด

3.7 การถดถอยและสหสัมพันธ์ (Regression and Correlation)

การศึกษาเกี่ยวกับความสัมพันธ์ของตัวแปรสุ่ม ตั้งแต่ 2 ตัวขึ้นไปที่เป็นปรากฏการณ์ทางอุทกนั้น วิธีหนึ่งที่สามารถใช้ในการวิเคราะห์ได้คือ ทฤษฎีการถดถอยและสหสัมพันธ์ การที่ตัวแปรตัวหนึ่งหรือชุดหนึ่งเปลี่ยนไป 1 หน่วย แล้วจะทำให้ตัวแปรอีกตัวหนึ่งหรือชุดหนึ่งเปลี่ยนไปเท่าใดนั้นเรียกว่า "การถดถอย" (regression) และ คำว่า "สหสัมพันธ์" (correlation) นั้นหมายถึงความสัมพันธ์ของตัวแปร ซึ่งจะเป็นตัวบอกให้ทราบว่า การคาดคะเนด้วยการถดถอยจะใช้ได้ดีเพียงใด ค่าที่ใช้ในการวัดเรียกว่า "สัมประสิทธิ์สหสัมพันธ์" (correlation coefficient) เช่น การกะประมาณค่าปริมาณฝนที่มีช่วงเวลาตั้ง 24 ชั่วโมงลงมา จากค่าปริมาณฝนรายวันนั้น วิธีหนึ่งที่อาจจะกระทำได้คือ การถดถอย ซึ่งค่าปริมาณฝนในช่วงเวลาที่ต้องการทราบคือตัวแปรตาม (dependent variable) นั้นจะมีค่าเปลี่ยนไปเท่าใด เมื่อค่าปริมาณฝนรายวันซึ่งเป็นตัวแปรอิสระ (independent variable) เปลี่ยนไป 1 หน่วย และกะประมาณค่าปริมาณฝนในช่วงเวลาที่ต้องการทราบจากค่าปริมาณฝนรายวันนี้ใช้ได้ดีเพียงใดจะขึ้นอยู่กับค่าสหสัมพันธ์ ซึ่งจะบอกได้ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ เป็นต้น

3.7.1 การถดถอยเชิงเส้นแบบง่าย (Simple linear regression)

ในการวิจัยครั้งนี้จะทำการศึกษาความสัมพันธ์ของตัวแปรครั้งละไม่เกิน 2 ตัว และจะเป็นการศึกษาแบบเชิงเส้นเท่านั้น ฉะนั้นในที่นี้จะกล่าวถึงเพียงการถดถอยเชิงเส้นแบบง่ายเท่านั้น ซึ่งมีหลักการโดยสังเขป คือ

เมื่อ x และ y เป็นตัวแปรที่มีความสัมพันธ์อยู่ในรูปคู่ลำดับ (x_i, y_i) สำหรับ $i = 1, 2, 3, \ldots, n$ ซึ่งถ้าให้ x_i เป็นตัวแปรอิสระกับ y_i เป็นตัวแปรตาม และมีความสัมพันธ์กันเป็นเส้นตรงแล้ว จะได้สมการถดถอยที่ประมาณจากกลุ่มตัวอย่างเช่น

$$\hat{y} = a+bx \qquad \ldots\ldots\ldots(3\text{-}101)$$

เมื่อ $\hat{y}$ คือค่าเฉลี่ยของการกะประมาณของตัวแปรตาม a และ b คือพารามิเตอร์ที่เรียกว่าสัมประสิทธิ์การถดถอย (regression coefficients)

และแต่ละคู่ลำดับนั้นจะมีความสัมพันธ์เป็น

$$y_i = a+bx_i+e_i \qquad \ldots\ldots\ldots(3\text{-}102)$$

เมื่อ e_i คือ residual ซึ่งมีค่าเท่ากับ $y_i-\hat{y}_i$ เมื่อ y_i คือค่าที่ได้จากการสังเกต (observed) ตามค่าของ x_i และ $\hat{y}_i$ คือค่าที่ได้จากเส้นถดถอย $\hat{y}_i = a+bx_i$

เส้นถดถอยที่ดีที่สุดนั้น คือเส้นที่จะทำให้ค่า $\sum_{i=1}^{n} e_i$ มีค่าน้อยที่สุด ซึ่งสามารถหาได้โดยวิธีกำลังสองน้อยที่สุด (method of least squares)

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i-(a+bx_i)]^2 \qquad \ldots\ldots\ldots(3\text{-}103)$$

minimized ค่า $\sum_{i=1}^{n} e_i^2$ เทียบกับค่า a และ b จะได้

$$\frac{\partial(\sum_{i=1}^{n} e_i^2)}{\partial a} = -2\sum_{i=1}^{n}(y_i-a-bx_i) = 0 \quad \ldots\ldots(3\text{-}104)$$

$$\frac{\partial(\sum_{i=1}^{n} e_i^2)}{\partial b} = -2\sum_{i=1}^{n}(y_i-a-bx_i)(x_i) = 0 \quad \ldots(3\text{-}105)$$

จาก (3-104) ; $$\sum_{i=1}^{n} y_i = na+b\sum_{i=1}^{n} x_i \quad \text{.......(3-106)}$$

จาก (3-105) ; $$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{.......(3-107)}$$

สมการที่ (3-106) และ (3-107) เรียกว่าสมการปกติ(normal equations)

ของสมการ $\hat{y} = a+bx$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \quad \text{.........(3-108)}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} \quad \text{........(3-109}$$

3.7.2 สหสัมพันธ์

การที่จะบอกว่าตัวแปรมีสหสัมพันธ์มากน้อยเพียงใดนั้น จะบอกด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งมีค่าอยู่ระหว่าง -1 ถึง 1 เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่า -1 หรือ 1 นั้นแสดงว่าตัวแปรมีความสัมพันธ์กันแน่นอน จุดพิกัดของค่า y_i ที่ได้จากการสังเกตของค่า x_i ทุกค่าจะอยู่บนเส้นถดถอย และถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าประมาณ -0.6 หรือ 0.6 แสดงว่าตัวแปรมีความสัมพันธ์กันไม่มากนัก และถ้าหากมีค่าเข้าใกล้ 0 ก็แสดงว่าตัวแปรนั้นไม่มีความสัมพันธ์กันเลย ค่าสัมประสิทธิ์สหสัมพันธ์นี้สามารถหาค่าได้จากสมการ

$$r = b\{[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2]/[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2]\}^{1/2} \quad \text{.....(3-110)}$$

เมื่อ r คือสัมประสิทธิ์สหสัมพันธ์

3.8 <u>การกะประมาณ</u> (Estimation)

การสรุปผลในเชิงสถิติเกี่ยวกับประชากร โดยทั่วไปนั้นจะอาศัยการสุ่มตัวอย่างมาหาค่าพารามิเตอร์ต่าง ๆ เพื่อใช้เป็นค่ากะประมาณของพารามิเตอร์ของประชากร ในกรณีที่กลุ่มประชากรมีขนาดใหญ่มาก หรือยังไม่สามารถทราบประชากรทั้งหมดได้ การประมาณค่าพารามิเตอร์นั้นจะแบ่งออกเป็น 2 ชนิดด้วยกันคือ การกะประมาณค่าแบบจุด (point estimation) และการกะประมาณแบบช่วง (interval estimation)

3.8.1 การกะประมาณแบบจุด

การกะประมาณแบบนี้เป็นการหาค่าสถิติต่าง ๆ จากกลุ่มตัวอย่างที่สุ่มมาจากประชากร เพื่อใช้เป็นค่ากะประมาณค่าพารามิเตอร์ต่าง ๆ ของกลุ่มประชากร เช่น การกะประมาณค่าเฉลี่ย $\bar{x}$ ของกลุ่มตัวอย่างที่สุ่มมาเพื่อใช้แทนค่าเฉลี่ยของประชากร μ เป็นต้น การกะประมาณแบบนี้ไม่สามารถจะคาดได้ว่าไม่มีความคลาดเคลื่อน เช่นไม่สามารถจะคาดได้ว่า $\bar{x}$ จะต้องเท่ากับ μ เพียงแต่คาดหวังว่า $\bar{x}$ จะต้องมีค่าไม่ห่างจาก μ มากนัก และในทำนองเดียวกันก็คาดหวังว่า S^2 จะมีค่าไม่ห่างไกลจาก σ^2 ด้วยเช่นกัน

การกะประมาณแบบจุดของพารามิเตอร์แต่ละตัวนั้น อาจจะมีวิธีการกะประมาณได้หลายวิธีด้วยกัน เช่น ค่าเฉลี่ยเลขคณิต, มัธยฐาน และฐานนิยม ของกลุ่มตัวอย่างต่างก็เป็นค่ากะประมาณของ μ เป็นต้น ฉะนั้นการที่เราจะใช้ค่ากะประมาณตัวใดแทนพารามิเตอร์ของประชากร ก็จะต้องตัดสินใจล่วงหน้าว่าจะใช้ตัวประมาณค่าตัวใด โดยพิจารณาจากคุณสมบัติของตัวกะประมาณว่า จะต้องเป็นตัวกะประมาณที่ไม่ลำเอียง (unbiased estimator) ที่มีประสิทธิภาพมากที่สุด ซึ่งสามารถเขียนเป็นสมการได้ว่า

$$\mu_{\hat{\theta}} = E(\theta) = \theta \qquad \ldots\ldots\ldots(3\text{-}111)$$

เมื่อ สถิติ $\hat{\theta}$ เป็นตัวกะประมาณค่าที่ไม่ลำเอียงของพารามิเตอร์ θ และ $E(\theta)$ คือค่าคาดคะเน (exspectation)

นิยามสำหรับการพิจารณาตัวกะประมาณที่ไม่ลำเอียงคือ "ในบรรดาตัวกะประมาณที่ไม่ลำเอียงทั้งหลายของพารามิเตอร์ θ ตัวกะประมาณที่มีความแปรปรวนน้อยที่สุดคือตัวกะประมาณที่มีประสิทธิภาพมากที่สุด" [จุฬาลงกรณ์มหาวิทยาลัย, คณาจารย์ภาควิชาคณิตศาสตร์(2526) ความน่าจะเป็นและสถิติ]

สำหรับวิธีการหาค่าสถิติต่าง ๆ ของกลุ่มตัวอย่างที่สุ่มมาจากประชากร เพื่อใช้เป็นค่ากะประมาณพารามิเตอร์ ของประชากรนั้น ได้กล่าวไว้แล้วในหัวข้อ 3.1.2

3.8.2 การกะประมาณแบบช่วง

เมื่อค่าพารามิเตอร์ θ ถูกกะประมาณโดยสถิติ $\hat{\theta}$ ซึ่งเป็นค่าที่หาได้จากกลุ่มตัวอย่างที่สุ่มมาแล้ว จะได้ว่า

$$P(U>\theta>L) \quad = \quad 1-\alpha \qquad \ldots\ldots\ldots(3\text{-}112)$$

เมื่อ U และ L คือค่าลิมิตแห่งความเชื่อมั่นบน (upper confidence limits) และลิมิตแห่งความเชื่อมั่นล่าง (lower confidence limits) ตามลำดับ, $1-\alpha$ เรียกว่าสัมประสิทธิ์แห่งความเชื่อมั่น (confidence coefficient) และ α คือระดับความมีนัยสำคัญ

จากสมการ (3-112) จะให้ความหมายว่า เมื่อมีการสุ่มตัวอย่างมาชุดหนึ่ง แล้วกำหนดเปอร์เซนต์ความเชื่อมั่น เท่ากับ $(1-\alpha)$ 100% เรียกช่วงที่สร้างขึ้นนี้ว่าช่วงแห่งความเชื่อมั่น (confidence interval) แล้ว ถ้ากระทำการหาช่วงด้วยวิธีเดียวกันซ้ำ ๆ 100 ครั้ง เราจะได้ช่วง 100 ชุด อาจจะมีค่าเท่ากันหรือไม่ก็ได้ และคาดว่าจะมี $(1-\alpha)$ 100 ชุด ซึ่งมีพารามิเตอร์ θ ที่ต้องการกะประมาณอยู่ในช่วงนั้น

การวิเคราะห์ความถี่และการถดถอย ดังที่ได้กล่าวมาแล้วในหัวข้อ 3.5 และ 3.7 นั้น เป็นการกะประมาณที่ค่าเฉลี่ยเท่านั้น ดังนั้นผลการวิเคราะห์อาจจะคลาดเคลื่อนไปจากค่าที่ได้จากการสังเกตอยู่บ้าง วิธีการที่จะวัดความเที่ยงตรง (accuracy) ของการวิเคราะห์นั้นสามารถหาได้ในรูปแบบต่าง ๆ เช่นการบอกถึงความเที่ยงตรงของผลการวิเคราะห์ด้วย ความคลาดเคลื่อนมาตราฐานของการกะประมาณ ที่ได้กล่าวมาแล้วในหัวข้อ 3.4 นั้น จะบอกให้ทราบว่าค่าของเหตุการณ์ที่เกิดขึ้นจะมีอยู่ 2 ใน 3 ส่วนที่อยู่ในช่วง $x_T \pm S_T$ แต่โดยทั่วไปนั้นนิยมสรุปผลในรูปแบบของลิมิตความเชื่อมั่น ซึ่งหาได้จากสมการ

$$U \quad = \quad x_T + t'_{(1-\alpha)} S_T \qquad \ldots\ldots\ldots(3\text{-}113)$$

$$L \quad = \quad x_T - t'_{(1-\alpha)} S_T \qquad \ldots\ldots\ldots(3\text{-}114)$$

สำหรับลิมิตความเชื่อมั่นของการวิเคราะห์ความถี่ ที่รอบปี T ปีใด ๆ

$$U \quad = \quad \hat{y} + t'_{(1-\alpha)} S_e \left[1 + \frac{1}{x_i} + (x_i - \bar{x})^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{1/2} \qquad \ldots\ldots(3\text{-}115)$$

$$L \quad = \quad \hat{y} - t'_{(1-\alpha)} S_e \left[1 + \frac{1}{x_i} + (x_i - \bar{x})^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{1/2} \qquad \ldots\ldots(3\text{-}116)$$

สำหรับลิมิตความเชื่อมั่นของการวิเคราะห์การถดถอย ที่ค่า x_i ใด ๆ (confidence interval on individual predicted)

เมื่อ $t'_{(1-\alpha)}$ คือค่าความเบี่ยงเบนมาตราฐานปกติ ที่ระดับความมีนัยสำคัญ $(1-\alpha)$ และ degree of freedom = n-2 และ S_e คือ ความคลาดเคลื่อนมาตราฐานของการกะประมาณของการวิเคราะห์การถดถอย

3.9 <u>ความถูกต้องของผลการวิเคราะห์</u>

ในการวิเคราะห์ความถี่ของข้อมูลทางอุทกวิทยา ความถูกต้องของผลการวิเคราะห์นั้นเป็นสิ่งหนึ่งที่จะต้องคำนึงถึง และการวิเคราะห์ความถี่นี้เป็นการวิเคราะห์เกี่ยวกับความน่าจะเป็น ดังนั้น ความถูกต้องในที่นี้จะหมายถึงความน่าเชื่อถือของผลที่ประเมินได้จากการวิเคราะห์นั้นเอง ซึ่งจะขึ้นอยู่กับองค์ประกอบต่าง ๆ คือ

3.9.1 จำนวนปีของข้อมูลที่ใช้ในการวิเคราะห์

ข้อมูลค่าปริมาณฝนสูงสุดที่ใช้ในการวิเคราะห์ ไม่ว่าจะเป็นแบบ annual series หรือ partial series นั้น เป็นการสุ่มตัวอย่าง เพื่อใช้แทนประชากรผลที่ได้จึงเป็นเพียงการกะประมาณ ดังได้กล่าวมาแล้วในหัวข้อ 3.8 ฉะนั้นขนาดของกลุ่มตัวอย่างย่อมมีความสำคัญต่อผลการวิเคราะห์ เพราะข้อมูลค่าปริมาณฝนสูงสุดนี้ไม่สามารถทราบประชากรได้ทั้งหมดและไม่สามารถจะทราบได้ว่ากลุ่มตัวอย่างที่สุ่มมานั้น ไม่มีความลำเอียง ดังนั้นถ้าจำนวนปีของข้อมูลที่ใช้ในการวิเคราะห์มีมากก็จะทำให้ผลการวิเคราะห์น่าเชื่อถือมากขึ้น เช่นถ้ามีจำนวนข้อมูลเพียง 10 ปี ย่อมจะทำให้มีความน่าเชื่อถือน้อยลงสำหรับการประเมินค่าที่รอบปีสูงถึง 25 ปี, 50 ปีหรือ 100 ปี

Hershfield & Wilson (1957) ได้แสดงให้เห็นถึงผลการวิเคราะห์ข้อมูลฝนของสถานีฝนที่ Hartford, Connecticut ว่าเมื่อมีข้อมูลเพิ่มขึ้นจาก 50 ปี เป็น 51 ปีนั้น เมื่อข้อมูลที่เพิ่มขึ้นอีก 1 ปีนั้นมีค่าสูงกว่าปีที่ผ่านมามาก ๆ จะทำให้ผลการวิเคราะห์ที่รอบปีเดียวกัน มีขนาดของเหตุการณ์แตกต่างกันมาก

สุรวุฒิ ประดิษฐานนท์ (2526) ได้ทำการวิเคราะห์ค่าปริมาณน้ำท่าสูงสุดที่ไหลเข้าเขื่อนอุบลรัตน์ เมื่อข้อมูลเปลี่ยนจาก 16 ปี (พ.ศ.2500-2520) ซึ่งไม่รวมค่าสูงสุดในปี 2521 ซึ่งทำให้เกิดวาตภัยครั้งใหญ่ไปเป็น 19 ปี (พ.ศ.2500-2523) นั้น ทำให้ความเหมาะสมเปรียบเทียบสูงสุดของการทดสอบความเหมาะสมของทฤษฎีการแจกแจงความน่าจะเป็น เปลี่ยนจากการ

แจกแจงแบบกัมเบลไปเป็น การแจกแจงแบบเพียร์สันชนิดที่ 3

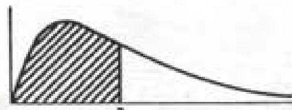
3.9.2 ทฤษฎีการแจกแจงที่ใช้ในการวิเคราะห์ความถี่

ตามหลักการของความน่าจะเป็นและสถิตินั้น ทฤษฎีการแจกแจงที่ใช้ในการวิเคราะห์ความถี่ที่เหมาะสมที่สุดนั้น จะให้ผลการวิเคราะห์ที่น่าเชื่อถือที่สุดและจะมีช่วงความเชื่อมั่นของการวิเคราะห์แคบที่สุด [Yevjevich (1972)] ฉะนั้นการทดสอบความเหมาะสมของทฤษฎีการแจกแจงดังกล่าวในหัวข้อ 3.6 จึงมีความจำเป็น

3.9.3 ความเที่ยงตรงของข้อมูล

ความผิดพลาดของการวิเคราะห์ความถี่นั้น อาจจะเกิด ได้เนื่องจากความคลาดเคลื่อนของข้อมูลค่าปริมาณฝน ซึ่งสำหรับการวิจัยในครั้งนี้ผู้วิจัยใช้วิธีอ่านข้อมูลกราฟฝนด้วยตาเปล่า ฉะนั้นสำหรับที่ช่วงเวลาสั้น ๆ ที่ 15 และ 30 นาที นั้น ข้อมูลอาจจะมีความคลาดเคลื่อนอยู่บ้างเล็กน้อย แต่สำหรับที่ช่วงเวลาเป็นชั่วโมงนั้นน่าจะมีความคลาดเคลื่อนน้อยกว่า แต่ทั้งนี้ขึ้นอยู่กับประสิทธิภาพของเครื่องวัดน้ำฝน และการติดตั้งกราฟฝน (pluviograph) ด้วย

ตารางที่ 3-1 ค่าเปอร์เซ็นไตล์ของการแจกแจงแบบไคสแคว (Percentile Values ($\chi^2_{\alpha,\nu}$) for the Chi-Square Distribution with Degrees of Freedom (Shaded area = α)) [Hann (1977)]



$\chi^2_{\alpha,\nu}$

ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

Source: Catherine M. Thompson, Table of percentage points of the χ^2 distribution, Biometrika, Vol. 32 (1941), by permission of the author and publisher.

ตาราง 3-2 ค่าวิกฤตสำหรับการทดสอบสถิติแบบ Kolmogorov-Smirnov (Critical Values for the Kolmogorov-Smirnov Test Statistic) [Hann (1977)]

Sample Size (n)	Significance Level .20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.829
4	.494	.525	.564	.624	.734
5	.446	.474	.510	.563	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.409	.486
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.391
17	.250	.266	.286	.318	.380
18.	.244	.259	.278	.309	.370
19	.237	.252	.272	.301	.361
20	.231	.246	.264	.294	.352
25	.21	.22	.24	.264	.32
30	.19	.20	.22	.242	.29
35	.18	.19	.21	.23	.27
40				.21	.25
50				.19	.23
60				.17	.21
70				.16	.19
80				.15	.18
90				.14	
100				.14	
Asymptotic Formula:	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Journal American Statistical Association 47:425-441, 1952, Z.W. Birnbaum.