บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

บทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับการพัฒนาเทคนิคการคัดเลือกตัวแบบสำหรับการวิเคราะห์ภัยแล้งในลุ่มน้ำปิงตอนบนได้แก่ ดรรชนีความแห้งแล้งของฝนที่ต่างจากค่าปรกติ เทคนิคการคัดเลือกตัวแบบ เกณฑ์การพิจารณาค่าความผิดพลาดแบบที่ 1 และอำนาจการทดสอบ การลงจุดตำแหน่ง การประมาณค่าพารามิเตอร์ การแจกแจงที่สนใจศึกษา รวมไปถึงงานวิจัยที่เกี่ยวข้อง

2.1 ดรรชนีความแห้งแล้งของฝนที่ต่างจากค่าปรกติ

ดรรชนีความแห้งแล้งของฝนที่ต่างจากค่าปรกติ (Standardized Precipitation Index: SPI) ได้พัฒนาขึ้นจากแนวคิดของ McKee และคณะ (Mckee et al, 1993 อ้างถึงใน กรมอุตุนิยมวิทยา, 2554) เพื่อเฝ้าดูสภาวะแห้งแล้งในช่วงเวลาต่างๆ ที่กำหนด โดยดูจากปริมาณฝนสะสมในแต่ละช่วงเวลาที่สนใจ ซึ่งอาจมีตั้งแต่ 1 เดือน 2 เดือน 3 เดือนจนถึง 72 เดือน โดยมีความเชื่อว่าปริมาณฝนโดยทั่วไปจะมีการกระจายในรูปแบบฟังก์ชันการแจกแจงแกมมา (Gamma Distribution) แต่เนื่องจากการศึกษาเพื่อหาค่าดรรชนี SPI จะต้องใช้ปริมาณฝนรวมเป็นหลักจึงได้พิจารณาโดยใช้ฟังก์ชันความหนาแน่นน่าจะเป็นสะสม (CDF) ของปริมาณฝนรวม แล้วทำการแปลง (Transform) ค่าดรรชนีให้เป็นค่าปรกติมาตรฐาน Z ซึ่งจะได้ค่าดรรชนี SPI ที่ต้องการแล้วนำมาจำแนกตามระดับความรุนแรงของภัยแล้งเพื่อบ่งชี้ถึงระดับความชุ่มชื้นและความแห้งแล้งของปริมาณฝนในแต่ละพื้นที่ ฟังก์ชันการแจกแจงแกมมาจะกำหนดโดยฟังก์ชันความหนาแน่นน่าจะเป็น (Probability Density Function : PDF) ดังนี้

$$f(x \mid \alpha, \beta) = \frac{x^{\alpha-1} e^{(-x/\beta)}}{\beta^{\alpha}\Gamma(\alpha)} \quad ; \alpha, \beta > 0 \; ; \; x > 0$$

โดยที่ $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ คือ ฟังก์ชันแกมมา

α แทน พารามิเตอร์บ่งรูปร่าง (Shape Parameter)

β แทน พารามิเตอร์บ่งขนาด (Scale Parameter)

ฟังก์ชันความหนาแน่นน่าจะเป็นสะสม (CDF) ของการแจกแจงแกมมา ดังนี้

$$F(x) = \frac{\Gamma_{x/\beta}(\alpha)}{\Gamma(\alpha)}$$

เมื่อ $\Gamma_{x/\sigma}(\alpha) = \int_0^{x/\sigma} t^{\alpha-1} e^{-t}\,dt$ คือ ฟังก์ชันแกมมาไม่สมบูรณ์ (Incomplete Gamma Function)

เนื่องจากแกมมาฟังก์ชันจะหาค่าไม่ได้เมื่อค่า $x = 0$ แต่โดยทั่วไปแล้วปริมาณฝนจะมีค่าเป็น 0 (ไม่มีรายงานฝนตก) ดังนั้นจึงต้องแปลงฟังก์ชันหนาแน่นน่าจะเป็นสะสม ดังนี้

$$H(x) = p + (1-p)G(x)$$

เมื่อ p คือ ความน่าจะเป็นที่จะไม่มีรายงานฝนตก ซึ่งมีค่าเท่ากับจำนวนวันที่ไม่มีฝนตก (m) หารด้วยจำนวนวันที่เราศึกษา (n) จากนั้นนำค่า H(x) มาแปลงเป็นค่าปรกติมาตรฐาน (Standard Normal) ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนมีค่าเท่ากับ 1 ซึ่งคือค่า SPI นั่นเอง ซึ่งมีสูตรการประมาณค่า Z หรือ SPI โดย Abramowitz และ Stegun (1965) ดังนี้

$$Z = SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \qquad \text{เมื่อ } 0 < H(x) \le 0.5$$

$$Z = SPI = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \qquad \text{เมื่อ } 0.5 < H(x) \le 1$$

โดยที่ $t = \sqrt{\ln\left(\frac{1}{(H(x))^2}\right)}$ เมื่อ $0 < H(x) \le 0.5$

$t = \sqrt{\ln\left(\frac{1}{(1-H(x))^2}\right)}$ เมื่อ $0.5 < H(x) \le 1$

เมื่อ $c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328$

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

ตารางที่ 2.1 เกณฑ์การแบ่งระดับความรุนแรงของดรรชนี SPI

ค่าดัชนี	**ระดับความรุนแรง**
มากกว่าหรือเท่ากับ 2	ฝนชุกมากที่สุด
1.50 ถึง 1.99	ฝนชุกมาก
1.00 ถึง 1.49	ฝนชุกปานกลาง
-0.99 ถึง 0.99	ฝนใกล้เคียงค่าปรกติ
-1.00 ถึง -1.49	ฝนแล้งปานกลาง
-1.50 ถึง -1.99	ฝนแล้งรุนแรง
น้อยกว่าหรือเท่ากับ -2	ฝนแล้งรุนแรงที่สุด

2.2 เทคนิคการคัดเลือกตัวแบบ

การวิเคราะห์เหตุการณ์ต่างๆทางอุทกวิทยา เช่น เหตุการณ์น้ำท่วมและเหตุการณ์ภัยแล้ง อาศัยลักษณะการแจกแจงของตัวแปรที่ใช้ในการศึกษาเป็นตัวช่วยในการวิเคราะห์ แม้ว่าตัวแปรที่ใช้ในการศึกษาในแต่ละพื้นที่จะเป็นตัวแปรตัวเดียวกัน แต่อาจมีลักษณะการแจกแจงที่แตกต่างกันในแต่ละพื้นที่ ฉะนั้นเทคนิคการคัดเลือกตัวแบบจึงเป็นเทคนิคที่จำเป็นในการวิเคราะห์เพื่อหาลักษณะการแจกแจงที่เหมาะสมสำหรับตัวแปรในแต่ละพื้นที่ ซึ่งในทางอุทกวิทยามีเทคนิคการคัดเลือกตัวแบบที่นิยมใช้หลากหลายเทคนิค สำหรับในการศึกษาครั้งนี้ มุ่งศึกษาเทคนิคการคัดเลือกตัวแบบต่างๆ ดังนี้

2.2.1 สถิติทดสอบสำหรับทดสอบภาวะสารูปดี

1) สถิติทดสอบแอนเดอร์สัน – ดาร์ลิ่ง (Anderson-Darling Test)

Anderson และ Darling (1954) ได้เสนอสถิติทดสอบที่ใช้ในการทดสอบภาวะสารูปดี เมื่อข้อมูลอยู่ในมาตรอันดับ (Ordinal Scale) และลักษณะการแจกแจงข้อมูลเป็นแบบต่อเนื่อง โดยใช้การวัดระยะห่างระหว่างฟังก์ชันการแจกแจงเชิงประจักษ์ $(F_n(x))$ กับการแจกแจงที่คาดหมาย $(F(x))$ ดังนี้

$$Q = n\int_{-\infty}^{\infty}[F_n(x) - F(x)]^2 \psi(x) dF(x)$$

เมื่อ $\psi(x)$ คือ ฟังก์ชันถ่วงน้ำหนัก (Weight Function)

$F(x)$ คือ ฟังก์ชันความหนาแน่นสะสม (CDF) ของการแจกแจงที่คาดหมาย

$F_n(x)$ คือ ฟังก์ชันการแจกแจงเชิงประจักษ์ (Empirical Distribution Function: EDF)

กรณีที่ $\psi(x)=1$ จะได้สถิติทดสอบคราแมร์-วอนมิส (Carmer-von Mises: CVM) ดังนี้

$$CVM = n\int_{-\infty}^{\infty}[F_n(x)-F(x)]^2\,dF(x)$$

กรณีที่ $\psi(x)=[F(x)\{1-F(x)\}]^{-1}$ จะได้สถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่ง ดังนี้

$$AD = n\int_{-\infty}^{\infty}\frac{[F_n(x)-F(x)]^2}{[F(x)\{1-F(x)\}]}\,dF(x)$$

$$AD = -n-\frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log F(x_i)+\log[1-F(x_{n+1-i})]]$$

$$= -n-\frac{1}{n}\sum_{i=1}^{n}[(2i-1)\log(F(x_i))+(2n+1-i)\log(1-F(x_i))]$$

2) สถิติทดสอบแอนเดอร์สัน – ดาร์ลิ่งที่ปรับปรุงของ Ahmad (Modified Anderson-Darling Test)

Ahmad และคณะ (1988) ได้ปรับปรุงสถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งโดยใช้ฟังก์ชันถ่วงน้ำหนักที่เน้นการเบี่ยงเบนที่หาง โดยแบ่งได้เป็น 2 กรณี คือ กรณีที่เน้นการเบี่ยงเบนที่หางทางขวา $\psi(x)=[1-F(x)]^{-1}$ จะได้สถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งที่ปรับปรุงของ Ahmad ดังนี้

$$AU = n\int_{-\infty}^{\infty}\frac{[F_n(x)-F(x)]^2}{1-F(x)}\,dF(x)$$

$$= \frac{n}{2}-2\sum_{i=1}^{n}F(x_i)-\sum_{i=1}^{n}\left\{2-\frac{2i-1}{n}\right\}\log\{1-F(x_i)\}$$

และกรณีที่เน้นการเบี่ยงเบนที่หางทางซ้าย $\psi(x)=[F(x)]^{-1}$ จะได้สถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งที่ปรับปรุง Ahmad ดังนี้

$$AL = n\int_{-\infty}^{\infty}\frac{[F_n(x)-F(x)]^2}{F(x)}\,dF(x)$$

$$= -\frac{3n}{2}+2\sum_{i=1}^{n}F(x_i)-\sum_{i=1}^{n}\frac{2i-1}{n}\log\{F(x_i)\}$$

2.2.2 เกณฑ์การทดสอบ

1) เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike Information Criterion: AIC)

Akaike (Akaike, 1973 อ้างถึงใน H. Bozdogan, 1987) ได้แสดงให้เห็นถึงความสัมพันธ์ระหว่างสารสนเทศคูลแบค-ไลเบลอร์และทฤษฎีภาวะความน่าจะเป็น (Likelihood Theory) โดยมีรายละเอียดดังต่อไปนี้

ให้ $\hat{I}(f,g) = C - E_f(\log(g(x\mid\theta)))$ เป็นค่าประมาณของฟังก์ชันการสูญเสียสารสนเทศคูลแบค-ไลเบลอร์ แต่เนื่องจากไม่ทราบค่าที่แท้จริงของ f ดังนั้นในการคัดเลือกตัวแบบ จึงต้องหาค่าต่ำสุดของ $E_f(\hat{I}(f,g)) = C - E_f(\log(g(x\mid\hat{\theta})))$ โดยที่ C เป็นค่าคงที่ ดังนั้นจะมีเพียงค่า $E_f(\log(g(x\mid\hat{\theta})))$ ซึ่งจะทำให้ได้ว่า

$$E_f(\log(g(x\mid\hat{\theta}))) = E_f E_x(\log(g(x\mid\hat{\theta}(y))))$$

เมื่อ $\hat{\theta}$ คือตัวประมาณค่าพารามิเตอร์ที่ได้จากวิธีภาวะความน่าจะเป็นสูงสุด (Maximum Likelihood Estimator: MLE) ของ θ ที่ได้จากข้อมูล y และการกำหนดฟังก์ชันการแจกแจง g

แต่อย่างไรก็ตาม $\hat{\theta}$ เป็นตัวประมาณค่าที่เอนเอียงของ $E_f E_x(\log(g(x\mid\hat{\theta}(y))))$ โดยค่าเอนเอียงนี้สามารถประมาณได้จากตัวแบบ g และกำหนดให้มีค่าเท่ากับ K ดังนั้น ตัวประมาณค่าที่ไม่เอนเอียงในการประมาณค่าของ $E_f E_x(\log(g(x\mid\hat{\theta}(y))))$ สำหรับตัวอย่างขนาดใหญ่และตัวแบบที่เหมาะสมของ $\log(L(\hat{\theta}\mid data)) - K$ อยู่ในรูป

$$\log(L(\hat{\theta}\mid data)) - K = C - \hat{E}_{\hat{\theta}}(I(f,\hat{g}))$$

เมื่อ $\hat{g} = g(.\mid\theta)$

หรือ $\hat{E}(K-L) = \log(L(\hat{\theta})) - K$

และเมื่อนำ -2 มาคูณสมการนี้โดยตลอด สมการที่ได้ใหม่จะมีชื่อเรียกว่าเกณฑ์สารสนเทศของอาไคเคะ (Akaike Information Criterion: AIC) นั่นคือ

$$AIC = -2\log(L(\hat{\theta})) + 2K$$

นอกจากนี้ค่า AIC อาจหาได้จากการประมาณค่าโดยวิธีกำลังสองน้อยที่สุด (Ordinary Leastsquare Method: OLS) ในกรณีนี้ค่า AIC จะอยู่ในรูป

$$AIC = n\log(\sigma^2) + 2K$$

$$\text{เมื่อ } \sigma^2 = \frac{\sum(\hat{\varepsilon}_i)^2}{n} \; ; i = 1,2,\ldots,n$$

โดยที่ $\hat{\varepsilon}_i$ คือ ค่าคลาดเคลื่อนที่ได้จากการประมาณตัวแบบที่เหมาะสมและ K เป็นผลรวมของจำนวนพารามิเตอร์ในตัวแบบที่รวมค่าคงที่และ σ^2

ดังนั้นจึงสรุปได้ว่า แนวคิดของการเลือกตัวแบบสำหรับวิธีการ AIC คือ การประมาณค่าตัวแบบ g ที่ทำให้ค่าคาดหมายสารสนเทศของคูลแบค-ไลเบลอร์ มีค่าต่ำที่สุด โดยจะมีความสัมพันธ์กับค่าภาวะความน่าจะเป็น (Likelihood)

2) เกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criterion: BIC)

Schwarz (Schwarz , 1978) ได้พัฒนาเกณฑ์การคัดเลือกตัวแบบที่ได้มาจากการดัดแปลงแบบเบส์ของเกณฑ์ AIC เรียกว่า เกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria : BIC) หรืออาจถูกเรียกว่า Schwarz Criterion หรือ Schwarz Information (SC) โดยมีสูตรดังนี้

$$BIC = -2\log L(\hat{\theta}) + K\log n$$

เมื่อ L คือ ฟังก์ชันภาวะความน่าจะเป็นสูงสุดของตัวแบบ

n คือ ขนาดตัวอย่าง

K คือ จำนวนพารามิเตอร์ของตัวแบบ

ดังนั้นจึงสรุปได้ว่า แนวคิดของ BIC ที่ใช้ในการคัดเลือกตัวแบบคือ จะทำการประมาณค่าตัวแบบ g ที่ทำให้มีค่าใกล้เคียงกับค่าสารสนเทศของคูลแบค-ไลเบลอร์ ที่สอดคล้องกับตัวแบบที่แท้จริง f โดยที่ BIC มีแนวโน้มในการเลือกตัวแบบที่ถูกต้อง แต่อาจจะไม่ใช่ตัวแบบที่เหมาะสมที่สุด โดยตัวแบบที่เลือกนั้นจะต้องมีค่า $I(f,g)$ และ K น้อยที่สุด แต่ค่าความน่าจะเป็นภายหลังของตัวแบบนั้นจะต้องมีค่ามากที่สุด

2.3 เกณฑ์การพิจารณาค่าความผิดพลาดแบบที่ 1 และอำนาจการทดสอบ

2.3.1 ประเภทของความผิดพลาดในการทดสอบภาวะสารูปดี

โดยทั่วไปในการทดสอบภาวะสารูปดีจะพบว่าผลการทดสอบที่ได้อาจมีความผิดพลาดเกิดขึ้นได้ 2 ลักษณะ คือ

1) ความผิดพลาดแบบที่ 1 (Type I Error)

เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานหลัก (H_0) เมื่อสมมติฐานหลัก (H_0) เป็นจริง ซึ่งความน่าจะเป็นที่จะเกิดความผิดพลาดในลักษณะนี้คือ ระดับนัยสำคัญ (α)

2) ความผิดพลาดแบบที่ 2 (Type II Error)

เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐานหลัก (H_0) เมื่อสมมติฐานหลัก (H_0) เป็นเท็จ

ตารางที่ 2.2 การเกิดความผิดพลาดในการทดสอบ

ข้อเท็จจริง	**การตัดสินใจ**	
	ยอมรับสมมติฐานหลัก (H_0)	**ปฏิเสธสมมติฐานหลัก (H_0)**
สมมติฐานหลัก (H_0) เป็นจริง	ตัดสินใจถูก	ความผิดพลาดแบบที่ 1
สมมติฐานหลัก (H_0) เป็นเท็จ	ความผิดพลาดแบบที่ 2	อำนาจการทดสอบ $(1-\beta)$

2.3.2 เกณฑ์การพิจารณาค่าความผิดพลาดแบบที่ 1

ในการศึกษาครั้งนี้ใช้เกณฑ์การพิจารณาค่าความผิดพลาดแบบที่ 1 ทั้งหมด 2 เกณฑ์ ได้แก่

1) เกณฑ์ของ Cochran (Cochran 1954)

กำหนดให้ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่เกิดจากการทดสอบเป็น $\hat{\alpha}$ ซึ่งสถิติทดสอบนั้นจะสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ก็ต่อเมื่อ $\hat{\alpha}$ มีค่าอยู่ในช่วง $[0.040\ ,\ 0.060]$ ที่ระดับนัยสำคัญ 0.05

2) เกณฑ์ของ Bradley (Bradley 1978)

กำหนดให้ค่าความน่าจะเป็นของความผิดพลาดแบบที่ 1 ที่เกิดจากการทดสอบเป็น $\hat{\alpha}$ ซึ่งสถิติทดสอบนั้นจะสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ก็ต่อเมื่อ $\hat{\alpha}$ มีค่าอยู่ในช่วง $[0.025\ ,\ 0.075]$ ที่ระดับนัยสำคัญ 0.05

ถ้าค่าความผิดพลาดแบบที่ 1 ที่ได้จากการทดลองนั้นมีค่าอยู่ในช่วงตามเกณฑ์ นั่นคือ สถิติทดสอบนั้นมีความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ได้ที่ระดับนัยสำคัญที่กำหนดไว้ โดยเกณฑ์ของ Cochran ใช้ในการพิจารณาความสามารถในการควบคุมความผิดพลาดแบบที่ 1 โดยยอมให้มีความคลาดเคลื่อนได้น้อย แต่ในเกณฑ์ของ Bradley จะยอมให้มีความคลาดเคลื่อนของความผิดพลาดแบบที่ 1 ได้มากกว่า ดังนั้นหากค่าความผิดพลาดแบบที่ 1 อยู่ในช่วงเกณฑ์ของ Cochran แสดงว่าความผิดพลาดแบบที่ 1 มีค่าใกล้เคียงกับระดับนัยสำคัญ แต่ถ้ามีค่าอยู่นอกช่วงเกณฑ์ของ Cochran แต่ยังอยู่ในช่วงเกณฑ์ของ Bradley แสดงว่าความผิดพลาดแบบที่ 1 ยังมีค่าพอรับได้ แต่ถ้าออกนอกช่วงเกณฑ์ของ Bradley แสดงว่าสถิติทดสอบไม่สามารถควบคุมค่าความผิดพลาดแบบที่1 ได้

2.3.3 อำนาจการทดสอบ

อำนาจการทดสอบ $(1-\beta)$ คือ ความน่าจะเป็นที่จะปฏิเสธสมมติฐานหลัก (H_0) เมื่อสมมติฐานหลัก (H_0) เป็นเท็จ สำหรับการศึกษาครั้งนี้จะพิจารณาอำนาจการทดสอบของสถิติทดสอบที่ความสามารถควบคุมความผิดพลาดแบบที่ 1 เมื่อใช้เกณฑ์ของ Cochran และ Bradley ได้เท่านั้น

2.4 การลงจุดตำแหน่ง

การลงจุดตำแหน่ง (Plotting Position) นิยมใช้อย่างกว้างขวางทางด้านวิศวกรรม อุทกวิทยา และทางด้านสถิติ โดยเฉพาะอย่างยิ่งการประมาณฟังก์ชันความน่าจะเป็น (Probability Plot) และสถิติอันดับ การลงจุดตำแหน่งมีสูตรทั่วไปดังนี้

$$P_i = \frac{i-a}{n+1-2a}$$

เมื่อ i แทน ลำดับที่ของข้อมูลเมื่อเรียงข้อมูลจากน้อยไปมาก

a แทน ค่าคงที่ มีค่าระหว่าง 0 ถึง 1

สำหรับค่า a เป็นค่าที่ถูกกำหนดขึ้นเพื่อหาตำแหน่งที่เหมาะสมที่สุดสำหรับข้อมูลที่มีการแจกแจงต่างกัน จากการศึกษาของ Ani Shabri (Ani Shabi, 2002) พบว่า มีผู้เสนอค่า a สำหรับแต่ละการแจกแจงหลายค่า เช่น Hazen เสนอให้ค่า $a = 0.50$ เหมาะสำหรับการแจกแจงค่าสุดขีด Weibull (Weibull, 1939) เสนอให้ค่า $a = 0$ เหมาะสำหรับทุกการแจกแจง Beard เสนอให้ค่า $a = 0.3175$ เหมาะสำหรับทุกการแจกแจง Blom เสนอให้ค่า $a = 0.375$ เหมาะสำหรับการแจกแจงปกติ Gringorten เสนอให้ค่า $a = 0.44$ เหมาะสำหรับการแจกแจงแบบเลขชี้กำลัง การแจกแจงกัมเบล และการแจกแจงค่าสุดขีดวางนัยทั่วไป Cunnane เสนอให้ค่า $a = 0.40$ เหมาะสำหรับการแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงเพียร์สันประเภทที่ 3

2.5 การประมาณค่าพารามิเตอร์

เนื่องจากการศึกษาครั้งนี้ เป็นการศึกษาโดยใช้การแจกแจงที่เป็นการแจกแจงเบ้ขวา ผู้วิจัยจึงสนใจศึกษาวิธีการประมาณพารามิเตอร์สำหรับข้อมูลที่มีการแจกแจงเบ้ขวาด้วยวิธี $L-moment$ (Hosking, 1990) ซึ่งเป็นการประมาณค่าพารามิเตอร์ที่ถูกพัฒนามาจากวิธีโมเมนต์ โดยใช้ความสัมพันธ์ของพารามิเตอร์ในรูปของฟังก์ชันเส้นตรง ดังนี้

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} EX_{r-k:r}\ , \quad r = 1,2,\ldots$$

ฉะนั้น จะได้ค่าประมาณ $L-moment$ ลำดับที่ 1 – 4 ดังนี้

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

เมื่อ $\beta_r = \int_0^1 x(F)F^r\,dF$

และจะได้ค่าประมาณ $L-moment\ ratios$ ดังนี้

$$\tau_2 = \lambda_2 / \lambda_1$$
$$\tau_3 = \lambda_3 / \lambda_2$$
$$\tau_4 = \lambda_4 / \lambda_2$$

เมื่อ τ_2 คือ อัตราส่วนไร้หน่วยของความแปรปรวนเชิงเส้น (L-coefficients of Variance : LCv)

τ_3 คือ อัตราส่วนไร้หน่วยของความเบ้เชิงเส้น (L-skewness : LCs)

τ_4 คือ อัตราส่วนไร้หน่วยของความโด่งเชิงเส้น (L-kurtosis : LCk)

2.6 การแจกแจงที่สนใจศึกษา

เนื่องจากการศึกษานี้เป็นการศึกษาเพื่อประยุกต์ใช้กับปริมาณฝนตามฤดูกาล ฉะนั้นผู้ศึกษาจึงสนใจศึกษาข้อมูลที่มีลักษณะการแจกแจงเบ้ขวา ดังนี้

2.6.1 การแจกแจงแกมมา (Gamma Distribution)

ถ้าตัวแปรที่สนใจศึกษาหรือเหตุการณ์ที่สนใจเกิดขึ้นในช่วงเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่ง ตัวแปรนั้นจะมีการแจกแจงปัวซง (Poisson Distribution) แต่หากตัวแปรที่สนใจศึกษานั้น คือระยะเวลาที่รอคอยจนกระทั่งเหตุการณ์ที่สนใจเหตุการณ์แรกเกิดขึ้น ตัวแปรนั้นจะมีการแจกแจงแบบเลขชี้กำลัง (Exponential Distribution) และหากตัวแปรที่สนใจศึกษานั้น คือระยะเวลาที่รอคอยจนกระทั่งได้จำนวนครั้งของเหตุการณ์ที่สนใจศึกษาครบตามต้องการแล้ว ตัวแปรนั้นจะมีการแจกแจงแกมมา

เมื่อตัวแปรสุ่ม X มาจากการแจกแจงแกมมา เขียนแทนด้วย $X \sim Gamma(\alpha, \beta)$ มีฟังก์ชันหนาแน่นน่าจะเป็น ดังนี้

$$f(x \mid \alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \quad , x > 0$$

เมื่อ α แทน พารามิเตอร์บ่งรูปร่าง (Shape Parameter) , $\alpha > 0$

β แทน พารามิเตอร์บ่งขนาด (Scale Parameter) , $\beta > 0$

มีฟังก์ชันการแจกแจงสะสมดังนี้

$$F(x) = \frac{\Gamma_{x/\beta}(\alpha)}{\Gamma(\alpha)}$$

เมื่อ $\Gamma_{x/\beta}(\alpha) = \int_0^{x/\beta} t^{\alpha-1} \exp(-t)dt \quad ; \alpha > 0$

2.6.2 การแจกแจงกัมเบล (Gumbel Distribution)

การแจกแจงกัมเบลเป็นการแจกแจงที่รู้จักกันในนามการแจกแจงค่าสุดขีดวางนัยทั่วไปแบบที่ 1 หรือบางครั้งอาจถูกเรียกว่าล็อกของการแจกแจงไวบูล หรือการแจกแจงเลขชี้กำลังคู่

เมื่อตัวแปรสุ่ม X มาจากการแจกแจงกัมเบลเขียนแทนด้วย $X \sim Gumbel(\beta,\xi)$ มีฟังก์ชันหนาแน่นน่าจะเป็นดังนี้

$$f(x\mid\beta,\xi)=\frac{1}{\beta}\exp\left[-\left(\frac{x-\xi}{\beta}\right)-\exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right],-\infty<x<\infty$$

เมื่อ β แทน พารามิเตอร์บ่งขนาด (Scale Parameter) , $\beta>0$

ξ แทน พารามิเตอร์บ่งตำแหน่ง (Location Parameter) $\xi>0$

มีฟังก์ชันการแจกแจงสะสม

$$F(x)=\exp\left[-\exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right]$$

2.6.3 การแจกแจงล็อกนอร์มอล (Lognormal Distribution)

เมื่อตัวแปรสุ่ม X มาจากการแจกแจงล็อกนอร์มอลเขียนแทนด้วย $X \sim Lognormal(\alpha,\beta,\xi)$ มีฟังก์ชันหนาแน่นน่าจะเป็น ดังนี้

$$f(x\mid\alpha,\beta,\xi)=\frac{1}{(x-\xi)\alpha\sqrt{2\pi}}\exp\left[-\frac{(\log(x-\xi)-\beta)^2}{2\alpha^2}\right]$$

เมื่อ α แทน พารามิเตอร์บ่งรูปร่าง (Shape Parameter) $\alpha>0$

β แทน พารามิเตอร์บ่งขนาด (Scale Parameter) $-\infty<\beta<+\infty$

ξ แทน พารามิเตอร์บ่งตำแหน่ง (Location Parameter) $0\le\xi<x$

มีฟังก์ชันการแจกแจงสะสมดังนี้

$$F(x)=\Phi\left(\frac{\log(x-\xi)-\beta}{\alpha}\right)$$

เมื่อ $\Phi(x)=\int_{-\infty}^{x}\left((2\pi)^{-1/2}\exp\left(-\frac{1}{2}x^2\right)\right)d(x)$

2.6.4 การแจกแจงค่าสุดขีดวางนัยทั่วไป (Generalized Extreme Value Distribution)

ถ้า X เป็นตัวแปรสุ่ม มาจากการแจกแจงค่าสุดขีดวางนัยทั่วไป เขียนแทนด้วย $X \sim GEV(\alpha, \beta, \xi)$ มีฟังก์ชันหนาแน่นน่าจะเป็น ดังนี้

$$f(x \mid \alpha, \beta, \xi) = \begin{cases} \exp - \left[1 + \alpha\left(\dfrac{x-\xi}{\beta}\right)\right]^{-1/\alpha} \dfrac{1}{\beta}\left(1 + \alpha\left(\dfrac{x-\xi}{\beta}\right)\right)^{(-1/\alpha)-1} & ;\ -\infty < x \le \xi - \dfrac{\beta}{\alpha} \quad for\, \alpha < 0 \\ & ;\ \xi - \dfrac{\beta}{\alpha} \le x < \infty \quad for\, \alpha > 0 \\ \exp\left[-\exp\left[-\left(\dfrac{x-\xi}{\beta}\right)\right]\dfrac{1}{\beta}\exp\left[-\left(\dfrac{x-\xi}{\beta}\right)\right]\right] & ;\ -\infty < x < \infty \quad for\, \alpha = 0 \end{cases}$$

มีฟังก์ชันการแจกแจงสะสมดังนี้

$$F(x) = \begin{cases} \exp\left[-\left(1 - \alpha\left(\dfrac{x-\xi}{\beta}\right)\right)^{1/\alpha}\right] & ;\ -\infty < x < \xi - \dfrac{\beta}{\alpha} \quad for\, \alpha < 0 \\ & ;\ \xi - \dfrac{\beta}{\alpha} \le x < \infty \quad for\, \alpha > 0 \\ \exp\left[-\exp\left[-\left(\dfrac{x-\xi}{\beta}\right)\right]\right] & ;\ -\infty < x < \infty \quad for\, \alpha = 0 \end{cases}$$

เมื่อ α แทน พารามิเตอร์บ่งรูปร่าง (Shape Parameter)

β แทน พารามิเตอร์บ่งขนาด (Scale Parameter)

ξ แทน พารามิเตอร์บ่งตำแหน่ง (Location Parameter)

สามารถเขียนฟังก์ชันการแจกแจงค่าสุดขีดวางนัยทั่วไปได้ 3 การแจกแจงตามลักษณะของพารามิเตอร์บ่งรูปร่าง (α) ได้แก่ เมื่อ $\alpha = 0$ เรียกว่า การแจกแจงกัมเบล (Gumbel Distribution) เมื่อ $\alpha > 0$ เรียกว่า การแจกแจงฟรีเชต (Frechet Distribution) และเมื่อ $\alpha < 0$ เรียกว่า การแจกแจงรีเวิร์สไวบูล (Reverse Weibull Distribution)

2.6.5 การแจกแจงรีเวิร์สไวบูล (Reverse Weibull Distribution)

การแจกแจงรีเวิร์สไวบูลเป็นการแจกแจงที่รู้จักกันในนามการแจกแจงค่าสุดขีดวางนัยทั่วไปแบบที่ 3 ซึ่งการแจกแจงนี้ถูกใช้กันอย่างแพร่หลายในงานด้านวิศวกรรม

เมื่อตัวแปรสุ่ม X มาจากการแจกแจงรีเวิร์สไวบูล เขียนแทนด้วย $X \sim \mathrm{Re}Wei(\alpha, \beta, \xi)$ มีฟังก์ชันหนาแน่นน่าจะเป็น ดังนี้

$$f(x \mid \alpha, \beta, \xi) = \frac{\alpha}{\beta}\left(\frac{x-\xi}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x-\xi}{\beta}\right)^{\alpha}\right]$$

3

เมื่อ α แทน พารามิเตอร์บ่งรูปร่าง (Shape Parameter) $,\alpha > 0$

β แทน พารามิเตอร์บ่งขนาด (Scale Parameter) $,\beta > 0$

ξ แทน พารามิเตอร์บ่งตำแหน่ง (Location Parameter) $\xi > 0$

มีฟังก์ชันการแจกแจงสะสม

$$F(t) = 1 - \exp\left[-\left(\frac{x-\xi}{\beta}\right)^{\alpha}\right]$$

2.6.6 การแจกแจงเพียร์สันประเภทที่ 3 (Pearson Type III Distribution)

เมื่อตัวแปรสุ่ม X มาจากการแจกแจงเพียร์สันประเภทที่ 3 เขียนแทนด้วย $X \sim PIII(\alpha, \beta, \xi)$ มีฟังก์ชันหนาแน่นน่าจะเป็นดังนี้

$$f(x \mid \alpha, \beta, \xi) = \frac{(x-\xi)^{\alpha-1} \exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]}{\beta^{\alpha}\Gamma(\alpha)}$$

เมื่อ α แทน พารามิเตอร์บ่งรูปร่าง (Shape Parameter)

β แทน พารามิเตอร์บ่งขนาด (Scale Parameter)

ξ แทน พารามิเตอร์บ่งตำแหน่ง (Location Parameter)

มีฟังก์ชันการแจกแจงสะสมดังนี้

$$F(x) = G\left(\alpha, \left(\frac{x-\xi}{\beta}\right)\right) / \Gamma(\alpha)$$

เมื่อ $G(\alpha, x) = \int_0^x t^{\alpha-1} \exp[-t] dt$ คือ ฟังก์ชันแกมมาไม่สมบูรณ์ (Incomplete Gamma Function)

2.7 งานวิจัยที่เกี่ยวข้อง

Laio, Baldassarre และ Montanari (2009)ได้ศึกษาเรื่อง เทคนิคการคัดเลือกตัวแบบสำหรับการวิเคราะห์ความถี่ของค่าสุดขีดทางอุทกวิทยา เพื่อศึกษาและเปรียบเทียบเทคนิคการคัดเลือกตัวแบบ 3 เทคนิค ได้แก่ Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) และ Anderson-Darling Test (AD) เมื่อข้อมูลที่ศึกษามีการแจกแจงกัมเบล การแจกแจงปรกติ การแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงเพียร์สันประเภทที่ 3 ด้วยขนาดตัวอย่างที่แตกต่างกัน และประยุกต์ใช้กับข้อมูลปริมาณน้ำท่าของประเทศอังกฤษ ผลการศึกษาสำหรับการจำลองข้อมูลพบว่า 1.) เทคนิคการคัดเลือกตัวแบบทั้งสามเทคนิคจะมีประสิทธิภาพเพิ่มขึ้นเมื่อข้อมูลที่ศึกษาเป็นข้อมูลที่มีการแจกแจง 2 พารามิเตอร์ ในทางกลับกันเทคนิคการคัดเลือกตัวแบบทั้งสามจะมีประสิทธิภาพลดลงเมื่อข้อมูลที่ศึกษาเป็นข้อมูลที่มีการแจกแจง 3 พารามิเตอร์ 2.) เกณฑ์ BICมีโอกาสเลือกการแจกแจงที่ถูกต้องมากกว่าการแจกแจงที่แท้จริง 3.) เกณฑ์ AICและเกณฑ์ BIC มักให้ผลการคัดเลือกการแจกแจงตรงกันในขณะที่สถิติทดสอบ ADให้ผลการคัดเลือกที่ต่างออกไป ส่วนผลการศึกษาสำหรับข้อมูลจริงพบว่า จากทั้ง 1,000 สถานีตรวจวัดปริมาณน้ำ พบว่า เกณฑ์ AIC และเกณฑ์ BIC มีการคัดเลือกการแจกแจงที่เหมาะสมตรงกันถึง 87% ในขณะที่สถิติทดสอบAD ให้ผลการคัดเลือกการแจกแจงที่เหมาะสมตรงกับเกณฑ์ BIC เพียง 51% และตรงกับเกณฑ์ AIC เพียง 49%

Heo และคณะ (2013) ได้ศึกษาเรื่อง การประมาณค่าสถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งที่ถูกปรับปรุงสำหรับการแจกแจงค่าสุดขีดเมื่อไม่ทราบค่าพารามิเตอร์บ่งตำแหน่ง โดยศึกษาอำนาจการทดสอบของสถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งที่ปรับปรุงของAhmad (MAD) เทียบกับ Chi-square Test (CS), Kolmogorov-Smirnov Test (KS), Cramer Von Mises Test (CVM) และ Anderson-Darling Test (AD) เมื่อข้อมูลที่ศึกษาเป็นข้อมูลที่มีการแจกแจงค่าสุดขีด ได้แก่ การแจกแจงล็อกกัมเบล การแจกแจงพาเรโตวางนัยทั่วไป การแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงโลจิสติกวางนัยทั่วไป และประยุกต์ใช้กับข้อมูลปริมาณฝนของเมืองนัมวาน เมืองแดวัลยอง เมืองอุลซัน และเมืองช็องอึบ ของประเทศเกาหลีใต้ ผลการศึกษาสำหรับการจำลองข้อมูลพบว่า สถิติทดสอบ MAD มีอำนาจการทดสอบสูงกว่า สถิติทดสอบ CS, สถิติทดสอบ KS, สถิติทดสอบ CVM และสถิติทดสอบ AD และยังพบว่าสถิติทดสอบ AD และสถิติทดสอบ MAD มีอำนาจการทดสอบสูงกว่าสถิติทดสอบ CS, สถิติทดสอบ KS และสถิติทดสอบ CVM ส่วนผลการศึกษาสำหรับข้อมูลจริง พบว่าสถิติทดสอบ AD และสถิติทดสอบ MAD ให้ผลการวิเคราะห์ไปในทางเดียวกันสำหรับข้อมูลปริมาณฝนของทุกเมืองในขณะที่สถิติทดสอบ CS, สถิติทดสอบ KS และสถิติทดสอบ CVM ให้ผลการวิเคราะห์ไปในทิศทางตรงกันข้ามกัน ฉะนั้นจึงเป็นการยากที่จะหาข้อสรุปเกี่ยวกับการแจกแจงที่เหมาะสมสำหรับข้อมูลปริมาณฝนของแต่ละเมือง

Shin, Jung, Jeong และ Heo (2011)ได้ศึกษาเรื่อง การประมาณค่าสถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งที่ถูกปรับปรุงสำหรับการแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงโลจิสติกวางนัยทั่วไป โดยศึกษาเกี่ยวกับอำนาจการทดสอบของสถิติทดสอบแอนเดอร์สัน-ดาร์ลิ่งที่ปรับปรุงของAhmad (MAD) เทียบกับ Chi-square Test (CS), Kolmogorov-Smirnov Test (KS) และ Cramer Von Mises Test (CVM) เมื่อข้อมูลที่ศึกษามีการแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงโลจิสติกวางนัยทั่วไป และประยุกต์ใช้กับข้อมูลปริมาณฝนของเมืองซกโซ และเมืองมิลยัง ในประเทศเกาหลีใต้ ผลการศึกษาสำหรับการจำลองข้อมูลพบว่า 1.) สถิติทดสอบ MAD มีอำนาจการทดสอบสูงกว่าสถิติทดสอบ CS สถิติทดสอบ KS และสถิติทดสอบ CVM เมื่อการแจกแจงของข้อมูลและการแจกแจงที่ศึกษาเป็นการแจกแจงเดียวกัน 2.) อำนาจการทดสอบของสถิติทดสอบ MAD จะเพิ่มขึ้นตามขนาดตัวอย่างเมื่อการแจกแจงของข้อมูลต่างจากการแจกแจงที่ศึกษา 3.) สถิติทดสอบ MAD เป็นสถิติทดสอบที่เหมาะสมสำหรับการหาการแจกแจงที่เหมาะของข้อมูลที่เป็นค่าสุดขีด ผลการศึกษาสำหรับข้อมูลจริงพบว่า การแจกแจงค่าสุดขีดวางนัยทั่วไปและการแจกแจงโลจิสติกวางนัยทั่วไปเป็นการแจกแจงที่เหมาะสมสำหรับข้อมูลปริมาณฝนของเมืองซกโซ แต่ไม่เหมาะสมสำหรับปริมาณฝนของเมืองมิลยัง

ดังนั้นผู้ศึกษาจึงสนใจศึกษาเกี่ยวกับการพัฒนาเทคนิคการคัดเลือกตัวแบบ โดยใช้ค่าการลงจุดตำแหน่งทั้งหมด 5 ค่า ได้แก่ ค่าการลงจุดตำแหน่งของ Weibull ค่าการลงจุดตำแหน่งของ Beard ค่าการลงจุดตำแหน่งของ Blom ค่าการลงจุดตำแหน่งของ Cunnane และค่าการลงจุดตำแหน่งของ Gringorten และประยุกต์ใช้ในการหาการแจกแจงที่เหมาะสมสำหรับข้อมูลปริมาณฝนตามฤดูกาลในลุ่มน้ำปิงตอนบน