# การเปรียบเทียบวิธีการแปลงข้อมูลของการแจกแจงแบบเลขชี้กำลังนัยทั่วไป และการประมาณปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่

# A Comparison of Data Transformation Methods of Generalized Exponential Distribution and Estimation of Summer Rainfall in Chiang Dao, Chiang Mai

ธีรพงษ์ แก้วประเสริฐ[1,2*] มานัดถุ์ คำกอง[1] และ พุฒิพงษ์ พุกกะมาน[1]

Theerapong Kaewprasert[1,2*], Manad Khamkong[1] and Putipong Bookkamana[1]

*[1]ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่*

*[2]บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่*

*[1]Department of Statistics, Faculty of Science, Chiang Mai University*

*[2]Graduate School, Chiang Mai University*

Received : 21 July 2017

Accepted : 26 September 2017

Published online : 4 October 2017

## บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลของการแจกแจงแบบเลขชี้กำลังนัยทั่วไปที่ขนาดตัวอย่างและพารามิเตอร์แตกต่างกันด้วยการแปลงแบบบ็อกซ์-ค็อกซ์ การแปลงแบบถอดรากที่สาม( $x^{1/3}$ ) และการแปลงแบบถอดรากที่สี่( $x^{1/4}$ ) ในกรณีที่สามารถแปลงข้อมูลเข้าใกล้การแจกแจงปรกติมากที่สุด โดยการเปรียบเทียบค่าร้อยละการยอมรับสมมุติฐานว่าง ($H_0$) ที่ว่าข้อมูลที่ทำการแปลงข้อมูลแล้วมีการแจกแจงปรกติ โดยการใช้เทคนิคมอนติคาร์โลจำนวน 10,000 ครั้งในแต่ละสถานการณ์ ผลการวิจัยพบว่าการแปลงแบบถอดรากที่สี่ให้ค่าร้อยละการยอมรับสมมุติฐานว่างมากที่สุด และนำไปประยุกต์ใช้ในการหาช่วงความเชื่อมั่นข้อมูลปริมาณฝนจากสถานีตรวจวัดปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่ โดยเฉลี่ยต่อปีมีค่าอยู่ระหว่าง 254.0344 มิลลิเมตร และ 340.1977 มิลลิเมตร ที่ระดับความเชื่อมั่น 95%

**คำสำคัญ** : การแจกแจงแบบเลขชี้กำลังนัยทั่วไป การแปลงแบบบ็อกซ์-ค็อกซ์ การแปลงแบบถอดรากที่สาม การแปลงแบบถอดรากที่สี่ ปริมาณฝนในฤดูร้อน

***Corresponding author. E-mail** : theerapong.nack@gmail.com

## Abstract

This study aims to compare the efficiency of the data transformation methods of generalized exponential distribution simulated with different sample size and parameters. Transformation methods including Box-Cox power transformation, Cube-Root transformation ( $x^{1/3}$ ) and Fourth-Root transformation ( $x^{1/4}$ ) improve to investigate method of data transformation can transform data to normalizing data. By comparing a percentage of accept the null hypothesis ($H_0$): the 10,000 repeated that data after transformation of the normal distribution are computed through the Monte Carlo technique. The simulation study's results show that Fourth-Root transformation gives a highest percentage of accept the null hypothesis. These methods apply to determine the confidence interval of rain in summer season at gauging station, Chiang Dao district, Chiang Mai, Thailand. Annually, the mean values are between 254.0344 mm. and 340.1977 mm. at 95% confidence level.


**Keywords** : generalized exponential distribution, Box-Cox transformation, Cube-Root transformation, Fourth-Root transformation, summer rainfall

**บทนำ**

การศึกษาข้อมูลเกี่ยวกับอายุการใช้งาน (Life time) ถูกนำมาใช้กับงานวิจัยในด้านต่าง ๆ อาทิ วิศวกรรมศาสตร์ การแพทย์ เป็นต้น ซึ่งลักษณะข้อมูลดังกล่าวสามารถจัดให้อยู่ในรูปแบบของการแจกแจงที่เหมาะสมได้หลายแบบด้วยกัน เช่น การแจกแจงแบบเลขชี้กำลัง (Exponential Distribution) การแจกแจงไวบูล (Weibull Distribution) การแจกแจงแกมมา (Gamma Distribution) เป็นต้น โดยการแจกแจงแบบเลขชี้กำลัง เป็นการศึกษาเวลารอคอยจนกระทั่งเกิดเหตุการณ์ที่สนใจศึกษาครั้งแรก เช่น ระยะเวลาที่รอคอยจนกระทั่งมีลูกค้าคนแรกเข้ามาใช้บริการในสำนักงาน ระยะเวลาที่รอคอยจนกระทั่งมีโทรศัพท์เข้ามาสายแรก ระยะเวลาการใช้งานของหลอดไฟจนกระทั่งหลอดไฟขาด เป็นต้น จากคุณสมบัติของการแจกแจงแบบเลขชี้กำลังดังกล่าวอาจมีข้อจำกัดก็คือเป็นการแจกแจงที่มีเพียงพารามิเตอร์เดียว คือ พารามิเตอร์แสดงขนาด (scale parameter) ซึ่งต่อมา Gupta และ Kundu (1999a) ได้เสนอการแจกแจงหนึ่งที่มีลักษณะการแจกแจงเบ้ขวาที่สามารถช่วยแก้ปัญหาในข้อจำกัดในการใช้งานของการแจกแจงแบบเลขชี้กำลัง นั่นคือการแจกแจงแบบเลขชี้กำลังนัยทั่วไป (Generalized Exponential Distribution : GE) ซึ่งเป็นการแจกแจงที่ประกอบด้วย 2 พารามิเตอร์ คือ พารามิเตอร์แสดงขนาด และพารามิเตอร์แสดงรูปร่าง (shape parameter) ซึ่งทำให้รูปแบบของการแจกแจงมีความสอดคล้องกับลักษณะข้อมูลที่หลากหลายยิ่งขึ้น และพบว่าข้อมูลปริมาณน้ำฝนเป็นข้อมูลที่สอดคล้องกับการแจกแจงแบบเลขชี้กำลังนัยทั่วไป ซึ่งถ้าหากต้องการนำข้อมูลนี้มาวิเคราะห์โดยใช้วิธีการทางสถิติที่มีข้อตกลงเบื้องต้นที่ว่าตัวอย่างสุ่มต้องมาจากประชากรที่มีการแจกแจงปรกติ (Normal distribution) วิธีการหนึ่งที่ใช้ในการแก้ปัญหา คือ การแปลงข้อมูล (Data transformation)

การแปลงข้อมูล คือ การแปลงค่าข้อมูลเดิมด้วยฟังก์ชันทางคณิตศาสตร์ เพื่อให้ค่าข้อมูลใหม่ที่ได้มีลักษณะการแจกแจงปรกติ ทำได้หลายวิธี เช่น การแปลงแบบลอการิทึม (Logarithm transformation) การแปลงแบบรากที่สอง (Square root transformation) การแปลงแบบการกลับเศษส่วน (Reciprocal transformation) การแปลงแบบเลขยกกำลัง (Power transformation) และการแปลงแบบบ็อกซ์-ค็อกซ์ (Box-Cox power transformation) เป็นต้น จากการศึกษาของ Nopparat (1999) พบว่า สำหรับการแจกแจงเบ้ขวา การแปลงแบบเลขยกกำลังจะมีความเหมาะสมเมื่อเลขยกกำลังมีค่าน้อยกว่า 1 และรูปแบบการแปลงที่มีค่าร้อยละของการยอมรับ $H_0$ การแจกแจงปรกติมากที่สุด ได้แก่ การแปลงแบบ

เลขยกกำลัง ในกรณีที่เลขยกกำลังมีค่าน้อยกว่า 1 รองลงมาคือ การแปลงแบบการกลับเศษส่วน การแปลงแบบลอการิทึม และการแปลงแบบรากที่สอง ตามลำดับ ต่อมา Jirawan (2011) พบว่าวิธีการแปลงข้อมูลที่มีการแจกแจงแบบเลขชี้กำลังให้มีการแจกแจงปรกติที่เหมาะสมที่สุดคือการแปลงแบบบ็อกซ์-ค็อกซ์ กรณีที่ค่าคงที่ของบ็อกซ์-ค็อกซ์มีค่าอยู่ระหว่าง 0 ถึง 1 และ Saithong (2004) พบว่าการทดสอบแอนเดอร์สัน-ดาร์ลิง (Anderson-Darling test : AD) มีกำลังการทดสอบสูงสุดสำหรับการทดสอบการแจกแจงแบบต่อเนื่องทั้งตัวอย่างขนาดใหญ่และตัวอย่างขนาดเล็ก รวมทั้ง Stephens (1974) พบว่าการทดสอบ AD มีกำลังการทดสอบดีกว่าการทดสอบคอลโมโกรอฟ (Kolmogorov-Smirnor test: KS) และการแปลงแบบถอดรากที่สามและรากที่สี่ Krishnamoorthy (2008) เสนอว่าเป็นวิธีการแปลงข้อมูลให้มีการแจกแจงปรกติที่เหมาะสมสำหรับการแจกแจงแกมมาซึ่งเป็นการแจกแจงเบ้ขวา

เนื่องจากการแจกแจงแบบเลขชี้กำลังนัยทั่วไปเป็นการแจกแจงเบ้ขวาและการแปลงแบบถอดรากที่สามและรากที่สี่สอดคล้องกับการแปลงแบบเลขยกกำลังในกรณีที่เลขยกกำลังมีค่าน้อยกว่า 1 ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพการแปลงแบบบ็อกซ์-ค็อกซ์ การแปลงแบบถอดรากที่สาม และการแปลงแบบถอดรากที่สี่ โดยใช้การทดสอบภาวะสารูปดี (Goodness of fit test) ของการทดสอบแอนเดอร์สัน-ดาร์ลิง รวมทั้งประยุกต์ใช้กับข้อมูลปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่

## วิธีดำเนินการวิจัย

### 1. วิธีการแปลงข้อมูล (Data transformation methods)

**1.1 การแปลงแบบบ็อกซ์-ค็อกซ์ (Box-Cox power transformation)** เสนอโดย Box และ Cox (1964) เป็นวิธีการแปลงข้อมูลเพื่อให้ข้อมูลที่ไม่มีการแจกแจงปรกติมีลักษณะข้อมูลเป็นการแจกแจงปรกติ โดยมีรูปแบบการแปลงดังนี้

$$y = \begin{cases} \dfrac{(x^{\beta} - 1)}{\beta} & ; \beta \neq 0,\ x > 0 \\ \ln(x) & ; \beta = 0,\ x > 0 \end{cases} \qquad (1)$$

เมื่อ $y$ คือ ค่าสังเกตหลังการแปลงข้อมูล

$x$ คือ ค่าสังเกตก่อนการแปลงข้อมูล

$\beta$ คือ ค่าคงที่ของบ็อกซ์-ค็อกซ์ ซึ่งสามารถหาได้จากการประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด

**1.2 การแปลงแบบถอดรากที่สาม (Cube-Root transformation)** เป็นวิธีการแปลงข้อมูลโดยการยกกำลังด้วยเศษหนึ่งส่วนสาม ซึ่งมีรูปแบบการแปลงดังนี้

$$y = x^{1/3} \qquad (2)$$

**1.3 การแปลงแบบถอดรากที่สี่** (Fourth-Root transformation) เป็นวิธีการแปลงข้อมูลโดยการยกกำลังด้วยเศษหนึ่งส่วนสี่ ซึ่งมีรูปแบบการแปลงดังนี้

$$y = x^{1/4} \tag{3}$$

**2. การแจกแจงที่เกี่ยวข้อง**

**2.1 การแจกแจงแบบเลขชี้กำลังนัยทั่วไป** (Generalized exponential distribution)

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบเลขชี้กำลัง ด้วยพารามิเตอร์แสดงขนาด $\lambda$ ที่มีฟังก์ชันการแจกแจงสะสม (Cumulative distribution function : cdf) ดังนี้

$$F(x) = 1 - e^{-\lambda x} \quad ; \ x > 0 , \lambda > 0 \tag{4}$$

เมื่อนำฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบเลขชี้กำลัง ยกกำลังด้วย $\alpha$ ซึ่งเป็นพารามิเตอร์แสดงรูปร่างจะได้ว่าตัวแปรสุ่ม X มีการแจกแจงแบบเลขชี้กำลังนัยทั่วไป โดยมีฟังก์ชันการแจกแจงสะสม คือ

$$F(x ; \alpha , \lambda) = \left(1 - e^{-\lambda x}\right)^{\alpha} \quad ; \ \alpha , \lambda , x > 0 \tag{5}$$

และฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function : pdf) คือ

$$f(x ; \alpha , \lambda) = \alpha\lambda \left(1 - e^{-\lambda x}\right)^{\alpha - 1} e^{-\lambda x} \quad ; \ \alpha , \lambda , x > 0 \tag{6}$$

ค่าคาดหมาย (expectation) $E(X) = \frac{1}{\lambda}\left(\psi(\alpha + 1) - \psi(1)\right)$

ความแปรปรวน (variance) $Var(X) = \frac{1}{\lambda^2}\left(\psi'(1) - \psi'(\alpha + 1)\right)$

โดยที่ $\psi(x) = \frac{d}{dx}\ln\left(\Gamma(x)\right) = \frac{\Gamma'(x)}{\Gamma(x)}$ และ $\psi'(x)$ คืออนุพันธ์ของ $\psi(x)$

**2.2 การแจกแจงปรกติ** (Normal distribution)

ถ้าตัวแปรสุ่ม X มีการแจกแจงปรกติด้วยค่าคาดหมาย $\mu$ และความแปรปรวน $\sigma^2$ โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็น คือ

$$f(x ; \mu , \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \ -\infty < x < \infty , -\infty < \mu < \infty , \sigma^2 > 0 \tag{7}$$

เมื่อ $\pi$ คือค่าคงที่เท่ากับ 3.1416... และ $e$ คือค่าคงที่เท่ากับ 2.7128...

**3. การทดสอบภาวะสารูปดี** (Goodness-of-fit test)

การทดสอบแอนเดอร์สัน-ดาร์ลิง เป็นการทดสอบที่ใช้ในการทดสอบภาวะสารูปดี ซึ่งเป็นการเปรียบเทียบระยะห่างระหว่างฟังก์ชันการแจกแจงสะสมของค่าสังเกตกับฟังก์ชันการแจกแจงสะสมที่คาดหมาย โดยที่ลักษณะการแจกแจงข้อมูลเป็นแบบต่อเนื่อง ซึ่ง Anderson และ Darling (1954) ได้เสนอไว้ดังนี้

$$AD = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\log u_j + \log(1-u_{n-j+1})\right] \quad (8)$$

เมื่อ $u_j = F(x_j)$ คือฟังก์ชันการแจกแจงสะสมที่คาดหมาย และ n คือขนาดตัวอย่าง

จะปฏิเสธ $H_0$ เมื่อ p-value ของค่าการทดสอบ AD มีค่าน้อยกว่า 0.05 (ระดับนัยสำคัญ 0.05)

**4. ข้อมูลที่ใช้ในการวิจัย**

**4.1 ข้อมูลจากการจำลอง**

4.1.1 จำลองข้อมูลให้มีการแจกแจงแบบเลขชี้กำลังนัยทั่วไปตามขนาดตัวอย่างและค่าพารามิเตอร์ที่กำหนด

1) กำหนดขนาดตัวอย่างที่ใช้ คือ n = 10, 20, 30, 50, 100

2) กำหนดค่าพารามิเตอร์ $\alpha$ = 0.5, 1.0, 1.5, 2.0, 2.5 และ $\lambda$ = 1

เนื่องจาก $\lambda$ เป็นพารามิเตอร์แสดงขนาด จึงไม่ส่งผลต่อลักษณะรูปร่างของการแจกแจง ในที่นี้จึง กำหนดให้ $\lambda$ = 1

4.1.2 ตรวจสอบว่าข้อมูลที่จำลองในขั้นตอนที่ 1 ว่ามีการแจกแจงแบบเลขชี้กำลังนัยทั่วไปหรือไม่ โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05

4.1.3 นำข้อมูลที่จำลองในแต่ละสถานการณ์มาทำการแปลงข้อมูลด้วยการแปลงแบบบ็อกซ์-ค็อกซ์ การแปลงแบบถอดรากที่สาม($x^{1/3}$) และการแปลงแบบถอดรากที่สี่($x^{1/4}$)

4.1.4 นำข้อมูลที่ได้ทำการแปลงข้อมูลแล้วมาทดสอบภาวะสารูปดีโดยการทดสอบของแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 ซึ่งสมมุติฐานการทดสอบ คือ

$H_0$ : ข้อมูลที่ทำการแปลงข้อมูลแล้วมีการแจกแจงปรกติ

$H_1$ : ข้อมูลที่ทำการแปลงข้อมูลแล้วไม่มีการแจกแจงปรกติ

4.1.5 ทำซ้ำขั้นตอนที่ 1-4 เป็นจำนวน 10,000 ครั้ง ในแต่ละสถานการณ์

4.1.6 คำนวณหาร้อยละการยอมรับ $H_0$

$$\text{ร้อยละการยอมรับ } H_0 = \frac{\text{จำนวนครั้งการยอมรับ } H_0}{10{,}000} \times 100\% \quad (9)$$

4.1.7 นำร้อยละการยอมรับ $H_0$ ที่ได้จากวิธีการแปลงข้อมูลทุกวิธีมาเปรียบเทียบกันในแต่ละสถานการณ์

**4.2 สำหรับข้อมูลจริง**

4.2.1 นำข้อมูลปริมาณฝนรายเดือนจากสถานีสำรวจปริมาณฝนของศูนย์อุทกวิทยาและบริหารน้ำภาคเหนือตอนบน กรมชลประทาน มาทำการสะสมเป็นปริมาณฝนตามฤดูกาล โดยที่ฤดูกาลของประเทศไทยแบ่งได้เป็น *3* ฤดูกาล ดังนี้

ฤดูร้อน ประมาณเดือนกุมภาพันธ์ถึงเดือนพฤษภาคม

ฤดูฝน ประมาณเดือนมิถุนายนถึงเดือนกันยายน

ฤดูหนาว ประมาณเดือนตุลาคมถึงเดือนมกราคมของปีถัดไป

ซึ่งในงานวิจัยนี้จะใช้ข้อมูลปริมาณฝนจากสถานีตรวจวัดปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่

4.2.2 ตรวจสอบว่าข้อมูลปริมาณฝนมีการแจกแจงแบบเลขชี้กำลังนัยทั่วไปหรือไม่โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05

4.2.3 ตรวจสอบว่าข้อมูลปริมาณฝนก่อนการแปลงข้อมูลมีการแจกแจงปรกติหรือไม่โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05

4.2.4 นำข้อมูลปริมาณฝนมาทำการแปลงข้อมูลด้วยวิธีการแปลงทั้ง 3 วิธี

4.2.5 ตรวจสอบว่าข้อมูลปริมาณฝนที่ทำการแปลงข้อมูลแล้วมีการแจกแจงปรกติหรือไม่ โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05

4.2.6 หาช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยข้อมูลปริมาณฝน( $\mu$ ) ที่ทำการแปลงข้อมูลแล้ว ดังนี้

$$P(\bar{y} - t_{0.975,\ n-1}S/\sqrt{n} \le \mu \le \bar{y} + t_{0.975,\ n-1}S/\sqrt{n}) = 0.95 \qquad (10)$$

เมื่อ $\bar{y}$ คือค่าเฉลี่ยของข้อมูลหลังการแปลงข้อมูล, S คือส่วนเบี่ยงเบนมาตรฐาน และ $t_{0.975,\ n-1}$ คือค่าของตัวแปรสุ่มทีที่ 0.975 และองศาเสรี (Degree of freedom) เท่ากับ n-1

**ผลการวิจัยและวิจารณ์ผล**

***ตารางที่ 1*** ร้อยละการยอมรับ $H_0$ เมื่อประชากรมีการแจกแจงแบบเลขชี้กำลังนัยทั่วไป $GE(\alpha,1)$ ที่ระดับนัยสำคัญ 0.05

| n | α | ร้อยละการยอมรับ $H_0$ | | |
|---|---|---|---|---|
| | | $x^{1/3}$ | $x^{1/4}$ | Box - Cox |
| 10 | 0.5 | 96.62% | 99.12% | 86.53% |
| | 1.0 | 98.69% | 99.72% | 95.21% |
| | 1.5 | 98.40% | 99.54% | 96.24% |
| | 2.0 | 97.76% | 99.22% | 96.79% |
| | 2.5 | 96.96% | 98.86% | 97.04% |
| 20 | 0.5 | 95.34% | 98.92% | 95.24% |
| | 1.0 | 98.91% | 99.67% | 97.71% |
| | 1.5 | 98.93% | 99.75% | 98.72% |
| | 2.0 | 98.37% | 99.75% | 98.29% |
| | 2.5 | 97.63% | 99.39% | 97.96% |
| 30 | 0.5 | 92.60% | 98.06% | 91.33% |
| | 1.0 | 98.61% | 99.29% | 94.03% |
| | 1.5 | 98.87% | 99.60% | 93.97% |
| | 2.0 | 98.56% | 99.72% | 94.14% |
| | 2.5 | 97.85% | 99.70% | 96.32% |
| 50 | 0.5 | 85.92% | 96.59% | 61.19% |
| | 1.0 | 97.90% | 98.48% | 74.58% |
| | 1.5 | 98.34% | 98.98% | 74.42% |
| | 2.0 | 98.05% | 99.43% | 76.13% |
| | 2.5 | 97.50% | 99.52% | 77.27% |
| 100 | 0.5 | 64.50% | 93.30% | 64.11% |
| | 1.0 | 95.87% | 97.50% | 89.82% |
| | 1.5 | 97.34% | 98.32% | 93.50% |
| | 2.0 | 96.90% | 99.05% | 89.50% |
| | 2.5 | 96.16% | 99.54% | 93.25% |

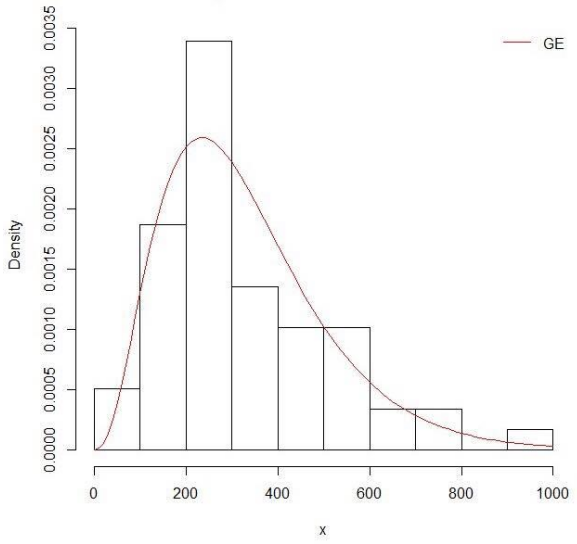
ผลการเปรียบเทียบประสิทธิภาพของวิธีการแปลงข้อมูล เมื่อพิจารณาตามขนาดตัวอย่างและค่าพารามิเตอร์ α พบว่าการแปลงแบบถอดรากที่สี่( $x^{1/4}$ ) ให้ค่าร้อยละการยอมรับสมมุติฐานมากที่สุด

เมื่อขนาดตัวอย่างเพิ่มขึ้นการแปลงแบบถอดรากที่สาม( $x^{1/3}$ ) ในกรณี $\alpha$ = 0.5, 1.0, 1.5 จะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มลดลง และ $\alpha$ = 2.0, 2.5 จะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มเพิ่มขึ้น ส่วนการ

แปลงแบบถอดรากที่สี่( $x^{1/4}$ ) ในกรณี $\alpha$ = 0.5, 1.0, 1.5, 2.0 จะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มลดลง และจะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มเพิ่มขึ้น เมื่อ $\alpha$ = 2.5 แต่การแปลงแบบบ็อกซ์-ค็อกซ์ จะให้ค่าร้อยละการยอมรับสมมุติฐานแตกต่างกันออกไปในแต่ละสถานการณ์

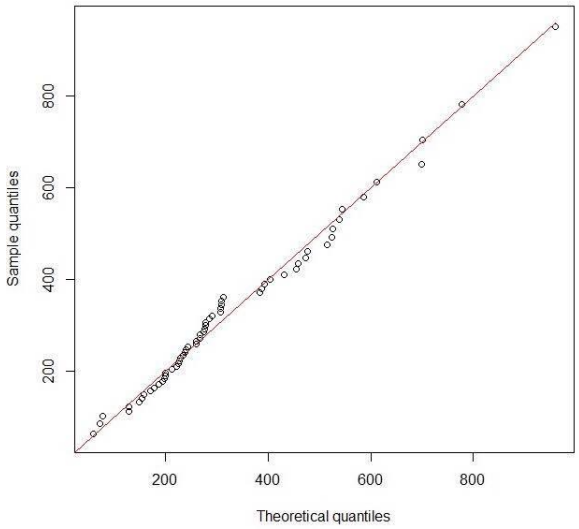
พิจารณาจากค่าพารามิเตอร์ $\alpha$ พบว่าการแปลงแบบถอดรากที่สาม( $x^{1/3}$ ) เมื่อ $\alpha$ = 0.5, 1.0, 1.5 จะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มเพิ่มขึ้นตามลำดับ และ $\alpha$ = 2.0, 2.5 จะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มลดลงตามลำดับเช่นกัน ส่วนการแปลงแบบถอดรากที่สี่( $x^{1/4}$ ) เมื่อ $\alpha$ เพิ่มขึ้น จะให้ค่าร้อยละการยอมรับสมมุติฐานมีแนวโน้มเพิ่มขึ้นด้วย และการแปลงแบบบ็อกซ์-ค็อกซ์ เมื่อ $\alpha$ เพิ่มขึ้น จะให้ค่าร้อยละการยอมรับสมมุติฐานแตกต่างกันออกไปในแต่ละสถานการณ์

สำหรับผลการวิเคราะห์ข้อมูลจริง กำหนดให้ $x$ คือข้อมูลปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่ ตั้งแต่ปี พ.ศ. 2500-2558 ซึ่งมีขนาดตัวอย่างเท่ากับ *59*(n = 59) ดังภาพที่ *1* โดยที่ ค่าเฉลี่ย( $\bar{x}$ ) เท่ากับ 329.4300 มิลลิเมตร(มม.) ส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 183.8327 มม. ค่าสูงสุดคือ 960.5000 มม. ค่าต่ำสุดคือ 58.9000 มม. ควอร์ไทล์ที่ 1 เท่ากับ 206.5500 มม. ควอร์ไทล์ที่ 2 เท่ากับ 275.2000 มม. ควอร์ไทล์ที่ 3 เท่ากับ 443.5000 มม.



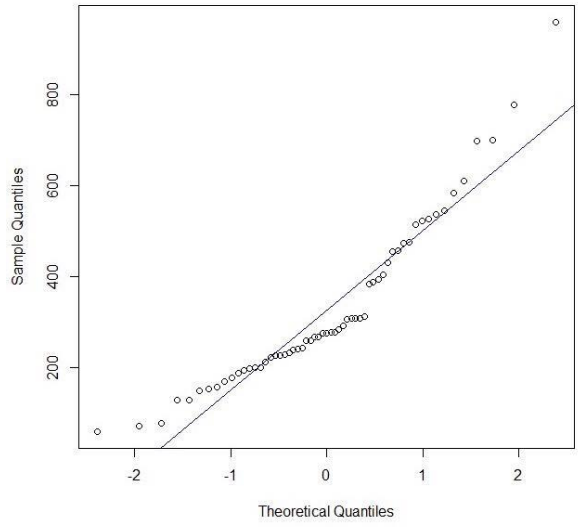

***ภาพที่ 1*** ปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่ ตั้งแต่ปี พ.ศ. 2500-2558 ด้วย pdf ของ GE

นำ $x$ มาทดสอบว่ามีการแจกแจงแบบเลขชี้กำลังนัยทั่วไป โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 พบว่า $x$ มีการแจกแจงแบบเลขชี้กำลังนัยทั่วไปจริง (p-value = 0.8218) โดยมีพารามิเตอร์แสดงรูปร่างเท่ากับ 4.3555 และพารามิเตอร์แสดงรูปร่างเท่ากับ 0.0066 ซึ่งแสดงด้วย Q-Q Plot ดังภาพที่ 2




***ภาพที่ 2*** Q-Q Plot ของ x สำหรับการแจกแจงแบบเลขชี้กำลังนัยทั่วไป

ทดสอบข้อมูล x ก่อนทำการแปลงข้อมูลว่ามีการแจกแจงปรกติหรือไม่ โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 พบว่า x ไม่มีการแจกแจงปรกติ (p-value = 0.0001) และแสดงด้วย Q-Q Plot ดังภาพที่ 3



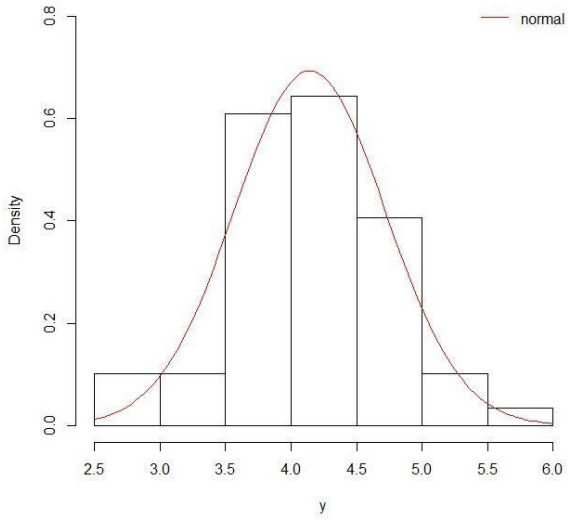

***ภาพที่ 3*** Q-Q Plot ของ x สำหรับการแจกแจงปรกติ

หลังจากนั้นนำข้อมูล x ทำการแปลงข้อมูลด้วยการแปลงข้อมูลทั้ง 3 วิธี พร้อมทั้งทดสอบภาวะสารูปดีได้ดังนี้

***ตารางที่ 2*** การทดสอบภาวะสารูปดีหลังทำการแปลงข้อมูลทั้ง 3 วิธี

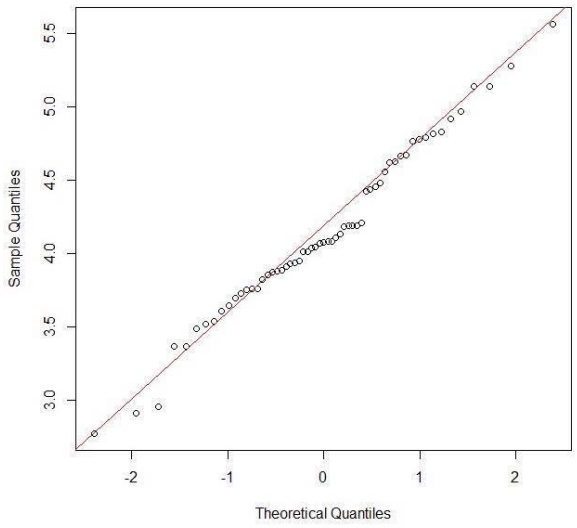
| Method | AD | p-value |
|---|---|---|
| $x^{1/3}$ | 0.5177 | 0.1814 |
| $x^{1/4}$ | 0.4556 | 0.2588 |
| Box - Cox | 0.9579 | 0.0145 |

ผลการทดสอบภาวะสารูปดีหลังทำการแปลงข้อมูลทั้ง 3 วิธี พบว่าการแปลง $x$ ด้วยการแปลงโดยใช้แบบถอดรากที่สี่( $x^{1/4}$ ) มีลักษณะข้อมูลเข้าใกล้การแจกแจงปรกติมากที่สุด เพราะให้ค่า AD น้อยที่สุด ดังตารางที่ 2 ซึ่งหมายความว่าระยะห่างระหว่างฟังก์ชันการแจกแจงสะสมของค่าสังเกตกับฟังก์ชันการแจกแจงสะสมที่คาดหมาย(การแจกแจงปรกติ) มีระยะห่างน้อยกว่าการแปลงวิธีอื่นๆ ซึ่งกำหนดให้ข้อมูลหลังการแปลงเป็นข้อมูล $y$ ดังภาพที่ 4




***ภาพที่ 4*** ข้อมูล $y$ หลังการแปลงข้อมูลด้วยการแปลงแบบถอดรากที่สี่( $x^{1/4}$ )

ทดสอบข้อมูล $y$ ว่ามีการแจกแจงปรกติหรือไม่ โดยใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 พบว่า $y$ มีการแจกแจงปรกติจริง (p-value = 0.2588) ดังตารางที่ 2 และแสดงด้วย Q-Q Plot ดังภาพที่ 5




***ภาพที่ 5*** Q-Q Plot ของ $y$ สำหรับการแจกแจงปรกติ

ทำการหาช่วงความเชื่อมั่นของ $y$ โดยใช้การแจกแจงที ที่ระดับนัยสำคัญ 0.05 เนื่องจากไม่ทราบค่าความแปรปรวนของประชากร โดยที่ ค่าเฉลี่ย $\bar{y}$ เท่ากับ 4.1435 และส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 0.5802 พบว่าที่ระดับความเชื่อมั่น 95% ข้อมูลเฉลี่ยของ $y$ มีค่าอยู่ระหว่าง 3.9923 (มม.)$^{1/4}$ และ 4.2947 (มม.)$^{1/4}$

ดังนั้นปริมาณฝนจากสถานีตรวจวัดปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่โดยเฉลี่ยมีค่าเท่ากับ $4.1435^4 = 294.7606$ มม. และช่วงความเชื่อมั่น 95% ของปริมาณฝนในฤดูร้อนของอำเภอเชียงดาวโดยเฉลี่ยต่อปีมีค่าอยู่ระหว่าง $3.9923^4 = 254.0344$ มม. และ $4.2947^4 = 340.1977$ มม. ซึ่งหมายความว่าจากข้อมูลปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว มีขนาดตัวอย่างเท่ากับ 59 นำมาสร้างช่วงความเชื่อมั่นได้ 100 ช่วง โอกาสที่จะพบว่าข้อมูลปริมาณฝนในฤดูร้อนของแต่ละปีของอำเภอเชียงดาว จะมีปริมาณฝนอยู่ในช่วงดังกล่าวแค่ 95 ช่วง

## สรุปผลการวิจัย

การแจกแจงแบบเลขชี้กำลังนัยทั่วไปมีข้อจำกัดในการประมาณค่าพารามิเตอร์และการหาช่วงความเชื่อมั่น แนวทางในการแก้ปัญหาวิธีหนึ่งคือการแปลงข้อมูล ซึ่งจากการเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลของการแจกแจงแบบเลขชี้กำลังนัยทั่วไปเพื่อให้ข้อมูลมีการแจกแจงปรกติ พบว่าการแปลงแบบถอดรากที่สี่( $x^{1/4}$ ) ให้ค่าร้อยละการยอมรับสมมุติฐานว่างมากที่สุดในทุกสถานการณ์ รองลงมาคือ การแปลงแบบถอดรากที่สี่( $x^{1/3}$ ) และการแปลงแบบบ็อกซ์-ค็อกซ์ ตามลำดับ ซึ่งสามารถนำการแปลงแบบถอดรากที่สี่( $x^{1/4}$ ) ไปใช้ในการหาช่วงความเชื่อมั่นสำหรับข้อมูลจริงได้ โดยการแปลงข้อมูลจริงซึ่งมีการแจกแจงแบบเลขชี้กำลังนัยทั่วไปเพื่อให้ข้อมูลนั้นมีการแจกแจงปรกติ( $y = x^{1/4}$ )และใช้ข้อมูล $y$ ในการหาช่วงความเชื่อมั่นแล้วอนุมานกลับไปหาข้อมูลจริงด้วย $x = y^4$ ซึ่งผู้วิจัยได้ประยุกต์ใช้กับข้อมูลจริงจากสถานีตรวจวัดปริมาณฝนในฤดูร้อนของอำเภอเชียงดาว จังหวัดเชียงใหม่

## กิตติกรรมประกาศ

งานวิจัยฉบับนี้เสร็จสมบูรณ์ได้เพราะได้รับความอนุเคราะห์จากบุคคลหลาย ๆ ท่าน รวมทั้งภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ และโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่ให้ทุนสนับสนุนการศึกษาและการทำวิจัย รวมถึงผู้ประเมินบทความทุกท่าน ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

## เอกสารอ้างอิง

Anderson, T. W., Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, *Vol. 49*, No. 268., pp. 765-769.

Box, G.E.P., Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, 211-252.

Gupta, R. D., Kundu, D. (1999a). Generalized Exponential Distributions. *Austral. N. Z. J. Statist*, *41*,173–188.

Jirawan, P. (2011). Data The Efficiency Comparisons Among the Box-Cox Power Transformation Method and the New Power Transformation Method for Exponential Distribution. Faculty of Science. Chonburi: Burapha University. (in Thai)

Krishnamoorthy, K., Mathew, T. and Mukherjee, S. (2008). Normal-Based Methods for a Gamma Distribution: Prediction and Tolerance Intervals and Stress-Strength Reliability. *American Statistical Association and the American Society for Quality*, *50*(1), 69-78

Nopparat, K. (1999). Data Transformation to Normal Distribution. Faculty of Commerce and Accountancy. Bangkok: Chulalongkorn University. (in Thai)

Saithong, J. (2004). A Comparison of Efficiency of the Test Statistics for Goodness of Fit Test. Faculty of Science. Nakhonpathom: Silpakorn University. (in Thai)

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, *69*, 730-737.