

การเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูล สำหรับข้อมูลที่มีการแจกแจงแกมมา

Efficiency Comparison of Data Transformation Methods for the Gamma Distributed Data

พิชชาพัส ภูรีจารุโรจน์*, วรวรรณ พัยคเวช, นัทธพงศ์ ตริสุธรรมมาศ,
ศุภานุช รักษาคม และจุฑาภรณ์ สิ้นสมบุญทอง
ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
แขวงลาดยาว เขตจตุจักร กรุงเทพมหานคร 10900

Phitchaphat Phureejarurot*, Worawan Payakwach, Nattapong Tresutummas,
Supanooch Raksakom and Juthaphorn Sinsomboonthong
Department of Statistics, Faculty of Science, Kasetsart University,
Ladyao, Chatuchak, Bangkok 10900

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลจากข้อมูลที่มีการแจกแจงแกมมา ให้มีการแจกแจงปกติ 4 วิธี คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบกำลัง วิธีการแปลงแบบรากที่สี่ และวิธีการแปลงแบบเลขชี้กำลังของ Manly โดยศึกษาจากข้อมูลจำลองที่มีการแจกแจงแกมมามีพารามิเตอร์สเกล (β) 1 และ 2 พารามิเตอร์รูปร่าง (α) 1, 3, 5, 10, 20, 50 และขนาดตัวอย่าง (n) ที่ศึกษา คือ 5, 10, 20, 30, 50, 100 กำหนดจำนวนครั้งการทำซ้ำในแต่ละสถานการณ์ 1,000 ครั้ง โดยเกณฑ์การเปรียบเทียบประสิทธิภาพพิจารณาจากร้อยละการยอมรับสมมติฐานหลักที่ว่าข้อมูลที่แปลงแล้วมีการแจกแจงปกติ โดยใช้สถิติทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 ถ้าวิธีใดมีค่าร้อยละการยอมรับสมมติฐานหลักสูงสุด แสดงว่าวิธีนั้นมีประสิทธิภาพดีที่สุด ผลการศึกษาพบว่าวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลังส่วนใหญ่ให้ประสิทธิภาพในเกือบทุกสถานการณ์ นอกจากนี้วิธีการแปลงแบบเลขชี้กำลังของ Manly ให้ประสิทธิภาพสูงสุดในกรณีตัวอย่างขนาดเล็ก ($n = 5, 10$) แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นพบว่าประสิทธิภาพการแปลงของวิธีการแปลงแบบเลขชี้กำลังของ Manly จะมีแนวโน้มลดลง

คำสำคัญ : การแปลงข้อมูล; วิธีการแปลงแบบบ็อกซ์-ค็อกซ์; วิธีการแปลงแบบกำลัง; วิธีการแปลงแบบรากที่สี่; วิธีการแปลงแบบเลขชี้กำลังของ Manly

Abstract

The objective of this research was to compare the efficiency of data transformation methods from Gamma distributed data to normally distributed data for four methods: Box-Cox transformation, power transformation, fourth root transformation, and exponential transformation of Manly methods. The simulation data generated from Gamma distribution with scale parameter (β) was equal to 1 and 2, shape parameter (α) was equal to 1, 3, 5, 10, 20, 50 and the sample size (n) of this study was equal to 5, 10, 20, 30, 50, and 100. Each situation was repeated 1,000 times. The criterion of efficiency comparison based on the acceptance percentage of a null hypothesis that the transformed data were normally distributed by using the Anderson-Darling test at the significance level 0.05. If the transformation method has a highest acceptance percentage of the null hypothesis, it means that this method is the most efficient method. According to this research, it was found that Box-Cox transformation and power transformation methods are the most efficient method for almost situations. In addition, exponential transformation of Manly method is the most efficient method for a small sample size ($n = 5, 10$). However, the efficiency of exponential transformation of Manly method tended to be decreased when the sample size increased.

Keywords: data transformation; Box-Cox transformation method; power transformation method; fourth root transformation method; exponential transformation of Manly method

1. บทนำ

การทำงานวิจัยสาขาต่าง ๆ ผู้วิจัยส่วนใหญ่จะนำเอาวิธีการทางสถิติมาใช้เพื่อวิเคราะห์หาข้อสรุปสำหรับงานวิจัยนั้น ๆ ซึ่งระเบียบวิธีทางสถิติแบบอิงพารามิเตอร์มักนิยมใช้กัน เนื่องจากมีประสิทธิภาพดี แต่ส่วนใหญ่จะมีข้อสมมติเบื้องต้นที่สำคัญ คือ ตัวอย่างสุ่มต้องมาจากประชากรที่มีการแจกแจงปรกติ เช่น การวิเคราะห์ความแปรปรวน แต่ความเป็นจริงข้อมูลที่ผู้วิจัยนำมาวิเคราะห์อาจไม่ได้มีการแจกแจงปรกติดังนั้นการใช้วิธีการแปลงข้อมูลจะช่วยให้ข้อมูลมีการแจกแจงปรกติ หรือมีการแจกแจงเข้าใกล้การแจกแจงปรกติมากที่สุด

การแปลงข้อมูลเป็นการปรับเปลี่ยนค่าของข้อมูลด้วยวิธีการทางคณิตศาสตร์ ซึ่งมีอยู่ด้วยกันหลาย

วิธี ได้แก่ วิธีการแปลงแบบลอการิทึม (logarithm transformation) วิธีการแปลงแบบรากที่สอง (square root transformation) วิธีการแปลงแบบเลขชี้กำลัง (exponential transformation) วิธีการแปลงแบบการกลับเศษส่วน (reciprocal transformation) วิธีการแปลงแบบกำลัง (power transformation) วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ (Box-Cox transformation) เป็นต้น งานวิจัยที่เกี่ยวข้องกับการศึกษาประสิทธิภาพของวิธีการแปลงข้อมูลจากข้อมูลที่มีการแจกแจงแบบต่าง ๆ ให้มีการแจกแจงปรกติส่วนใหญ่ใช้เกณฑ์ในการเปรียบเทียบประสิทธิภาพโดยพิจารณาจากร้อยละการยอมรับสมมติฐานหลัก ดังงานวิจัยต่อไปนี้ การศึกษาของ Krataithong [1] พบว่าสำหรับการแจกแจงเบ้ขวา วิธีการแปลงข้อมูลที่มีประสิทธิภาพดีที่สุด ได้แก่ วิธีการ

แปลงแบบกำลัง กรณีที่เลขกำลังมีค่าน้อยกว่า 1 รongลงมา คือ วิธีการแปลงแบบการกลับเศษส่วน วิธีการแปลงแบบลอการิทึม และวิธีการแปลงแบบรากที่สอง ตามลำดับ ต่อมา Ngamjarus [2] ได้ศึกษาหาวิธีการแปลงข้อมูลที่สามารถแปลงข้อมูลที่มีการแจกแจงแบบเลขชี้กำลัง การแจกแจงแกมมา การแจกแจงโคกำลังสอง การแจกแจงไวบูล และการแจกแจงลิอานอร์มอลให้มีการแจกแจงปกติ โดยพิจารณาวิธีการแปลงข้อมูลทั้งหมด 4 วิธี คือ วิธีการแปลงแบบกำลัง วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงข้อมูลลอการิทึม และวิธีการแปลงข้อมูลรากที่สอง ผลการวิจัยพบว่าการแปลงข้อมูลที่มีการแจกแจงแบบเลขชี้กำลัง การแจกแจงแกมมา การแจกแจงโคกำลังสอง และการแจกแจงไวบูล วิธีการแปลงข้อมูลที่เหมาะสมที่สุด คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ สำหรับการแจกแจงลิอานอร์มอล วิธีที่เหมาะสมที่สุด คือ วิธีการแปลงแบบลอการิทึม ต่อมา Kerdawang [3] ได้ศึกษาและเปรียบเทียบวิธีการแปลงข้อมูลให้มีการแจกแจงปกติ โดยพิจารณาวิธีการแปลงข้อมูลทั้งหมด 4 วิธี คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบตัดแปลงของบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบเลขชี้กำลังของ Manly (exponential transformation of Manly) และวิธีการแปลงแบบกำลังแบบใหม่ของ In-Kwon-Yeo สรุปได้ว่ากรณีที่ข้อมูลมีค่าบวกและค่าลบ สำหรับการแจกแจงแบบเบ้ขวาและการแจกแจงแบบเบ้ซ้าย ที่ความเบ้ระดับต่ำ วิธีการแปลงแบบตัดแปลงของบ็อกซ์-ค็อกซ์ให้ร้อยละการยอมรับสมมติฐานหลักสูงที่สุดเป็นส่วนใหญ่ ส่วนที่ความเบ้ระดับสูงขึ้น วิธีการแปลงแบบเลขชี้กำลังให้ร้อยละการยอมรับสมมติฐานหลักสูงที่สุดเป็นส่วนใหญ่ ส่วนกรณีที่ข้อมูลมีค่าบวก สำหรับการแจกแจงแบบเบ้ขวา ในทุกระดับความเบ้ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ให้ร้อยละการยอมรับสมมติฐานหลักสูงที่สุดเป็นส่วนใหญ่ สำหรับการแจกแจงแบบเบ้ซ้าย

ในทุกระดับความเบ้ วิธีการแปลงแบบเลขชี้กำลังให้ร้อยละการยอมรับสมมติฐานหลักสูงที่สุดเป็นส่วนใหญ่ ต่อมาในปี พ.ศ. 2554 Paiboolworachat [4] ได้เปรียบเทียบประสิทธิภาพระหว่างวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลังแบบใหม่ ภายใต้การแจกแจงแบบเลขชี้กำลัง ผลการวิจัยพบว่าวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ เมื่อเลขกำลังของการแปลงมีค่าอยู่ระหว่าง 0 ถึง 1 จะมีร้อยละการยอมรับสมมติฐานหลักมากที่สุด เมื่อเลขกำลังของการแปลงเท่ากับ 1 พบว่าวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลังแบบใหม่จะมีร้อยละการยอมรับสมมติฐานหลักเท่ากัน เมื่อเลขกำลังของการแปลงมากกว่า 1 พบว่าวิธีการแปลงแบบกำลังแบบใหม่จะมีร้อยละการยอมรับสมมติฐานหลักมากที่สุด ซึ่งในปีเดียวกัน Chortirat [5] ได้ศึกษาและเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูล 4 วิธี คือ วิธีการแปลงแบบฟังก์ชันความคลาดเคลื่อน (error function transformation) วิธีการแปลงแบบ dual power (dual power transformation) วิธีการแปลงแบบเลขชี้กำลังของ Manly และวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ ภายใต้ข้อมูลที่มีการแจกแจงแบบไวบูล พบว่าในตัวอย่างขนาดเล็กและขนาดกลาง วิธีการแปลงข้อมูลแบบเลขชี้กำลังของ Manly มีประสิทธิภาพดีที่สุด และในตัวอย่างขนาดใหญ่ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์มีประสิทธิภาพดีที่สุด นอกจากนี้ในปี พ.ศ. 2560 Kaewprasert และคณะ [6] ได้เปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลของการแจกแจงแบบเลขชี้กำลังนัยทั่วไปที่ขนาดตัวอย่างและพารามิเตอร์แตกต่างกัน 3 วิธี คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบรากที่สาม และวิธีการแปลงแบบรากที่สองในกรณีที่สามสามารถแปลงข้อมูลเข้าใกล้การแจกแจงปกติมากที่สุด ผลการวิจัยพบว่าวิธีการแปลงแบบรากที่สองให้ร้อยละการยอมรับสมมติฐานหลักมากที่สุด

ส่วนใหญ่การวิเคราะห์ข้อมูลทางสถิติแบบอิงพารามิเตอร์นั้น มีข้อสมมติเบื้องต้นที่สำคัญ คือ ตัวอย่างสุ่มต้องมาจากประชากรที่มีการแจกแจงปกติ ผู้วิจัยจำเป็นต้องทราบรูปแบบการแจกแจงของข้อมูลชุดนั้น ๆ โดยทดสอบการแจกแจงของข้อมูลก่อนนำมาวิเคราะห์ ซึ่งตัวสถิติทดสอบมีอยู่หลายวิธี ได้แก่ ตัวสถิติทดสอบแอนเดอร์สัน-ดาร์ลิง (Anderson-Darling statistic) ตัวสถิติทดสอบชาปิโร-วิลค์ (Shapiro-Wilk statistic) เป็นต้น การศึกษาของ Channarong [7] ได้เปรียบเทียบการทดสอบการแจกแจงปกติด้วยตัวสถิติทดสอบ 4 วิธี คือ ตัวสถิติทดสอบ Q ตัวสถิติทดสอบ D ตัวสถิติทดสอบคอลโมโกรอฟ-สมิรโนฟแบบ two stage delta-corrected และตัวสถิติทดสอบแอนเดอร์สัน-ดาร์ลิง โดยแบ่งการศึกษาเป็น 2 ส่วน คือ ส่วนที่ 1 เปรียบเทียบความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ส่วนที่ 2 เปรียบเทียบกำลังการทดสอบ ผลการศึกษาในส่วนที่ 1 พบว่าตัวสถิติทดสอบแอนเดอร์สัน-ดาร์ลิงเป็นตัวสถิติทดสอบที่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ดีที่สุดในทุกระดับนัยสำคัญและทุกขนาดตัวอย่าง ส่วนผลการศึกษาในส่วนที่ 2 พบว่าเมื่อกลุ่มตัวอย่างขนาดเล็ก ตัวสถิติทดสอบ D และตัวสถิติทดสอบแอนเดอร์สัน-ดาร์ลิง มีกำลังการทดสอบสูงใกล้เคียงกัน สำหรับตัวอย่างขนาดกลางและขนาดใหญ่ ตัวสถิติทดสอบ Q มีกำลังการทดสอบสูงที่สุด ต่อมา Pinthongpan [8] ศึกษาความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และเปรียบเทียบกำลังการทดสอบของสถิติทดสอบ 5 วิธี ได้แก่ สถิติทดสอบของเบต้าโดยใช้พื้นฐานการวิเคราะห์การถดถอยเชิงพหุ สถิติทดสอบแอนเดอร์สัน-ดาร์ลิงที่ปรับปรุงตามอัตราส่วนภาวะน่าจะเป็นสถิติทดสอบแอนเดอร์สัน-ดาร์ลิง สถิติทดสอบชาปิโร-วิลค์ และสถิติทดสอบของชาปิโร-ฟรานเซีย (Shapiro-Francia statistic) พบว่าสถิติทดสอบแอนเดอร์สัน-

ดาร์ลิงที่ปรับปรุงตามอัตราส่วนภาวะน่าจะเป็นและสถิติทดสอบแอนเดอร์สัน-ดาร์ลิง สามารถควบคุมความผิดพลาดแบบที่ 1 ในทุกขนาดตัวอย่าง ขณะที่สถิติทดสอบชาปิโร-ฟรานเซียไม่สามารถควบคุมความผิดพลาดแบบที่ 1 ในทุกขนาดตัวอย่าง สำหรับกำลังการทดสอบพบว่าส่วนใหญ่สถิติทดสอบแอนเดอร์สัน-ดาร์ลิงที่ปรับปรุงตามอัตราส่วนภาวะน่าจะเป็น มีกำลังการทดสอบสูงที่สุด รองลงมาสถิติทดสอบแอนเดอร์สัน-ดาร์ลิง ส่วนสถิติทดสอบของเบต้าโดยใช้พื้นฐานการวิเคราะห์การถดถอยเชิงพหุ และสถิติทดสอบชาปิโร-วิลค์มีกำลังการทดสอบที่ไม่แตกต่างกัน

งานวิจัยที่เกี่ยวข้องซึ่งกล่าวมาข้างต้น ส่วนใหญ่จะศึกษาวิธีการแปลงข้อมูลในการแจกแจงที่เบ้ซ้าย เบ้ขวา หรือการแจกแจงอื่น ๆ ที่หลากหลาย แต่ไม่ได้ศึกษารณิพารามิเตอร์แตกต่างกันที่หลากหลายในการแจกแจงหนึ่ง ๆ เพื่อคุณลักษณะค่าพารามิเตอร์ที่แตกต่างกันมีผลกับประสิทธิภาพของวิธีการแปลงแต่ละวิธีหรือไม่ ดังนั้นในงานวิจัยนี้จึงสนใจศึกษาข้อมูลที่มีการแจกแจงแกมมาเท่านั้นโดยศึกษาค่าพารามิเตอร์สเกลและพารามิเตอร์รูปร่างที่แตกต่างกันตั้งแต่ค่าน้อยไปจนถึงมาก เนื่องจากบ่อยครั้งพบว่าในงานวิจัยข้อมูลที่เกี่ยวข้องรวบรวมมาได้อาจมีการแจกแจงแกมมาแต่ผู้วิจัยมีความจำเป็นต้องแปลงข้อมูลก่อนเพื่อให้สอดคล้องกับข้อสมมติเบื้องต้นจึงจะนำข้อมูลนั้นไปวิเคราะห์สถิติที่ต้องการได้ เช่น ข้อมูลเกี่ยวกับอายุการใช้งานของอุปกรณ์อิเล็กทรอนิกส์ในกระบวนการควบคุมคุณภาพ ซึ่งผู้วิจัยอาจมีความจำเป็นต้องใช้ข้อมูลดังกล่าวไปสร้างแผนภูมิควบคุมคุณภาพซึ่งข้อสมมติเบื้องต้นของการสร้างแผนภูมิตั้งกล่าวส่วนใหญ่มักสร้างมาจากพื้นฐานที่ว่าข้อมูลต้องมีการแจกแจงปกติ ได้แก่ แผนภูมิควบคุมชีวฮาร์ท แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบเลขชี้กำลัง [9] เป็นต้น ดังนั้นจึงต้องมีการแปลงข้อมูลก่อนจะนำไปสร้างแผนภูมิควบคุม

คุณภาพดังกล่าวข้างต้นนี้ได้ ซึ่งในการวิจัยครั้งนี้ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลสำหรับข้อมูลที่มีการแจกแจงแกมมาโดยมีพารามิเตอร์สเกลแตกต่างกัน 2 ระดับ และพารามิเตอร์รูปร่างของการแจกแจงแตกต่างกันตั้งแต่ค่าน้อยไปจนถึงค่ามากทั้งหมด 4 วิธี คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบกำลัง วิธีการแปลงแบบรากที่สี่ และวิธีการแปลงแบบเลขชี้กำลังของ Manly โดยใช้สถิติทดสอบแอนเดอร์สัน-ดาร์ลิงสำหรับทดสอบภาวะสารูปดี เพื่อศึกษาวิธีการแปลงข้อมูลวิธีใดมีประสิทธิภาพมากกว่ากัน ในสถานการณ์ที่แตกต่างกัน นอกจากนี้งานวิจัยนี้ยังได้จำลองข้อมูลเพื่อศึกษาและสรุปผลการวิจัยเพื่อหาค่าเลขกำลังของการแปลงที่เหมาะสมที่สุดสำหรับวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบกำลัง และวิธีการแปลงแบบเลขชี้กำลังของ Manly โดยศึกษาค่าเลขกำลังของการแปลงมีค่าเท่ากับ -1 ถึง 1 ด้วยอัตราเพิ่มขึ้นทีละ 0.1 สำหรับสถานการณ์ต่าง ๆ เพื่อเป็นแนวทางในการเลือกวิธีการแปลงข้อมูลพร้อมทั้งการเลือกใช้ค่าเลขกำลังที่เหมาะสมที่สุดสำหรับข้อมูลที่มีการแจกแจงแกมมาในสถานการณ์ที่แตกต่างกันได้อย่างมีประสิทธิภาพให้แก่ผู้วิจัยที่สนใจศึกษาข้อมูลในงานวิจัยด้านต่าง ๆ ต่อไป

2. วิธีการวิจัย

การศึกษาเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลสำหรับข้อมูลที่มีการแจกแจงแกมมาที่ขนาดตัวอย่างและพารามิเตอร์แตกต่างกันด้วยการแปลงข้อมูล 4 วิธี คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ (Box-Cox transformation, BC) วิธีการแปลงแบบรากที่สี่ (fourth root transformation, FR) วิธีการแปลงแบบกำลัง (power transformation, PT) และวิธีการแปลงแบบเลขชี้กำลังของ Manly (exponential transformation of Manly, ET) มีขั้นตอนการวิจัยดังนี้

2.1 จำลองข้อมูลด้วยโปรแกรม SAS เวอร์ชัน 9.4 ให้มีการแจกแจงแกมมาตามขนาดตัวอย่างและค่าพารามิเตอร์ที่กำหนด ดังนี้

2.1.1 กำหนดค่าพารามิเตอร์สเกล (β) เท่ากับ 1 และ 2 และค่าพารามิเตอร์รูปร่าง (α) เท่ากับ 1, 3, 5, 10, 20 และ 50

2.1.2 กำหนดขนาดตัวอย่าง (n) เท่ากับ 5, 10, 20, 30, 50 และ 100

2.2 กำหนดกำลังของวิธีการแปลงข้อมูล (p) สำหรับวิธีการแปลงแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบกำลัง และวิธีการแปลงแบบเลขชี้กำลังของ Manly มีค่าเท่ากับ -1 ถึง 1 ด้วยอัตราเพิ่มขึ้นทีละ 0.1

2.3 นำข้อมูลที่จำลองในแต่ละสถานการณ์ในข้อ 2.1 มาแปลงข้อมูลด้วยการแปลงข้อมูลทั้ง 4 วิธี ดังนี้

2.3.1 วิธีการแปลงข้อมูลแบบบ็อกซ์-ค็อกซ์ เป็นวิธีการแปลงข้อมูลโดยใช้เลขยกกำลังอีกรูปแบบหนึ่งที่ได้ปรับปรุงโดยความร่วมมือของนักสถิติ จอร์จ บ็อกซ์ และ เซอร์ เดวิด ร็อบบี ค็อกซ์ ในปี ค.ศ. 1964 มีรูปแบบการแปลง [5] ดังนี้

$$y_{BC}(x) = \begin{cases} \frac{(x^p - 1)}{p} ; p \neq 0, x > 0 \\ \log_{10}(x) ; p = 0, x > 0 \end{cases} \quad (1)$$

เมื่อ $y_{BC}(x)$ คือ ข้อมูลหลังการแปลงด้วยวิธีการแปลงแบบบ็อกซ์-ค็อกซ์; x คือ ข้อมูลก่อนการแปลง; p คือ กำลังของการแปลง

2.3.2 วิธีการแปลงข้อมูลแบบรากที่สี่เป็นวิธีการแปลงข้อมูลโดยการยกกำลังด้วยเศษหนึ่งส่วนสี่ มีรูปแบบการแปลง [6] ดังนี้

$$y_{FR}(x) = x^{\frac{1}{4}} \quad (2)$$

เมื่อ $y_{FR}(x)$ คือ ข้อมูลหลังการแปลงด้วยวิธีการแปลงแบบรากที่สี่; x คือ ข้อมูลก่อนการแปลง

2.3.3 วิธีการแปลงข้อมูลแบบกำลังเป็นการแปลงข้อมูลโดยการยกกำลังด้วย p มีรูปแบบการแปลง [1] ดังนี้

$$y_{PT}(x) = x^p \tag{3}$$

เมื่อ $y_{PT}(x)$ คือ ข้อมูลหลังการแปลงด้วยวิธีการแปลงแบบกำลัง; x คือ ข้อมูลก่อนการแปลง; p คือ กำลังของการแปลงข้อมูล

2.3.4 วิธีการแปลงข้อมูลแบบเลขชี้กำลังของ Manly เป็นวิธีการแปลงข้อมูลที่สามารถใช้ได้กับข้อมูลที่มีทั้งค่าบวกและค่าลบ มีรูปแบบการแปลง [5] ดังนี้

$$y_{ET}(x) = \begin{cases} [\exp(px) - 1] / p & ; p \neq 0 \\ x & ; p = 0 \end{cases} \tag{4}$$

เมื่อ $y_{ET}(x)$ คือ ข้อมูลหลังการแปลงด้วยวิธีการแปลงแบบเลขชี้กำลังของ Manly; x คือ ข้อมูลก่อนการแปลง; p คือ กำลังของการแปลงข้อมูล

2.4 นำข้อมูลหลังการแปลงมาทดสอบการแจกแจงปกติ โดยการใช้การทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 โดยสมมติฐานการทดสอบ คือ

H_0 : ข้อมูลมีการแจกแจงปกติ

H_1 : ข้อมูลไม่มีการแจกแจงปกติ

ตัวสถิติทดสอบ [10,11] คือ

$$AD^* = AD \left(1 + \frac{4}{n} - \frac{25}{n^2} \right) \tag{5}$$

เมื่อ $AD = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln z_i + \ln(1 - z_{n+1-i})]$ (6)

โดย $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ แทนข้อมูลที่เรียงลำดับจาก

น้อยไปมาก n แทนขนาดตัวอย่าง $z_i = \phi \left(\frac{x_{(i)} - \bar{x}}{s} \right)$

โดย $x_{(i)}$ แทนค่าสถิติอันดับที่ i ; $i = 1, 2, 3, \dots, n$;

$\phi(\cdot)$ แทนฟังก์ชันการแจกแจงความน่าจะเป็นสะสมของการแจกแจงปกติมาตรฐาน; \bar{x} เป็นค่าเฉลี่ยเลข

คณิตของข้อมูลตัวอย่าง; s เป็นค่าเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่าง

จะปฏิเสธสมมติฐานหลัก เมื่อค่า AD^* ที่คำนวณได้ตามสมการ (5) มีค่ามากกว่า 0.751 ซึ่งเป็นค่าวิกฤติ ณ ระดับนัยสำคัญที่กำหนด

2.5 ทำซ้ำขั้นตอนที่ 2.1 ถึง 2.4 เป็นจำนวน 1,000 ครั้ง

2.6 คำนวณหาร้อยละการยอมรับสมมติฐาน H_0 ให้ A คือ จำนวนครั้งการยอมรับสมมติฐาน H_0

$$\text{ร้อยละการยอมรับ } H_0 = \frac{A}{1,000} \times 100 \tag{7}$$

2.7 นำร้อยละการยอมรับสมมติฐาน H_0 ที่ได้จากการแปลงข้อมูลทุกวิธีมาเปรียบเทียบกับในแต่ละสถานการณ์

2.8 สรุปผลการวิจัยในแต่ละสถานการณ์

3. ผลการวิจัย

งานวิจัยได้เปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูล สำหรับข้อมูลที่มีการแจกแจงแกมมา โดยศึกษาวิธีการแปลงข้อมูล 4 วิธี คือ วิธีการแปลงแบบบ็อกซ์-ค็อกซ์ (วิธี BC) วิธีการแปลงแบบรากที่สี่ (วิธี FR) วิธีการแปลงแบบกำลัง (วิธี PT) วิธีการแปลงแบบเลขชี้กำลังของ Manly (วิธี ET) โดยจำลองข้อมูลที่มีการแจกแจงแกมมา ด้วยค่าพารามิเตอร์รูปร่าง (α) 1, 3, 5, 10, 20 และ 50 และค่าพารามิเตอร์สเกล (β) 1, 2 ภายใต้ขนาดตัวอย่างที่ศึกษา ได้แก่ 5, 10, 20, 30, 50 และ 100 กำหนดจำนวนครั้งที่ทำซ้ำแต่ละสถานการณ์ 1,000 ครั้ง และเปรียบเทียบประสิทธิภาพด้วยค่าร้อยละการยอมรับสมมติฐานหลักที่ว่าข้อมูลที่แปลงแล้วมีการแจกแจงปกติ ผลการวิจัยแสดงในตารางที่ 1 ถึง 2 และรูปที่ 1 ถึง 2

Table 1 Percentages of acceptance a null hypothesis for Gamma distribution data with $\beta = 1$ and different values of α

α	Sample sizes (n)	Data transformation methods			
		FR method	PT method	BC method	ET method
1	5	95.4	95.8 (p=0.3)*	95.8 (p=0.3)*	94.6 (p=-0.7)
	10	94	95.2 (p=0.3)*	95.2 (p=0.3)*	90.6 (p=-0.9)
	20	93.2	94.4 (p=0.3)*	94.4 (p=0.3)*	83.7 (p=-0.8)
	30	94.1	95.9 (p=0.3)*	95.9 (p=0.3)*	73.4 (p=-0.8)
	50	93.2	94.6 (p=0.3)*	94.6 (p=0.3)*	49.8 (p=-0.8)
	100	91.3	94.3 (p=0.3)*	94.3 (p=0.3)*	7.9 (p=-0.8)
3	5	96	96 (p=0.4)	96 (p=0.4)	96.5 (p=-0.3)*
	10	94.7	95.3 (p=0.4)	95.3 (p=0.4)	96.2 (p=-0.3)*
	20	95.1	95.6 (p=0.3)	95.6 (p=0.3)	95.9 (p=-0.2)*
	30	94.9	95.7 (p=0.4)*	95.7 (p=0.4)*	94.7 (p=-0.2)
	50	94.3	94.8 (p=0.3)*	94.8 (p=0.3)*	91.7 (p=-0.2)
	100	93.8	94.8 (p=0.3)*	94.8 (p=0.3)*	83.4 (p=-0.2)
5	5	95.4	95.5 (p=0.4)	95.5 (p=0.4)	95.8 (p=-0.1)*
	10	95.2	95.1 (p=0.4, 0.5)	95.1 (p=0.4, 0.5)	95.8 (p=-0.2)*
	20	96.1	96.2 (p=0.3)*	96.2 (p=0.3)*	95.9 (p=-0.1, -0.2)
	30	95	95.3 (p=0.4)*	95.3 (p=0.4)*	94.4 (p=-0.1)
	50	94.8	95 (p=0.3)*	95 (p=0.3)*	92.5 (p=-0.1)
	100	93.4	94.8 (p=0.4)*	94.8 (p=0.4)*	86.7 (p=-0.1)
10	5	95.1	95.4 (p=0.7)*	95.4 (p=0.7)*	95.1 (p=-0.1)
	10	95.9	95.8 (p=0.2)	95.8 (p=0.2)	96 (p=-0.1)*
	20	95.4	95.5 (p=0.3)*	95.5 (p=0.3)*	95.3 (p=-0.1)
	30	94.8	95.3 (p=0.4)*	95.3 (p=0.4)*	94.1 (p=-0.1)
	50	94.9	95.5 (p=0.4)*	95.5 (p=0.4)*	92.1 (p=-0.1)
	100	94.5	94.8 (p=0.3)*	94.8 (p=0.3)*	89.3 (p=-0.1)
20	5	95.3	95.5 (p=0.4, 0.5)*	95.5 (p=0.4, 0.5)*	94.7 (p=0)
	10	95.7	95.9 (p=0.4, 0.5)*	95.9 (p=0.4, 0.5)*	93.9 (p=0)
	20	95.3	95.4 (p=0.2)*	95.4 (p=0.2)*	92.1 (p=0)
	30	95.3	95.4 (p=0.3)*	95.4 (p=0.3)*	90.2 (p=0)
	50	94.5	95.1 (p=0.4)*	95.1 (p=0.4)*	86.1 (p=0)
	100	94.2	94.3 (p=0.3)*	94.3 (p=0.3)*	74.4 (p=0)
50	5	94.7	94.9 (p=0.7, 0.9)*	94.9 (p=0.7, 0.9)*	94.8 (p=0)
	10	95.2	95.6 (p=0.5, 0.6)*	95.6 (p=0.5, 0.6)*	94.6 (p=0)
	20	95.2	95.4 (p=0.4)*	95.4 (p=0.4)*	93.1 (p=0)
	30	94.6	94.7 (p=0.3)*	94.7 (p=0.3)*	92.6 (p=0)
	50	94.6	95.2 (p=0.6)*	95.2 (p=0.6)*	92.8 (p=0)
	100	94.4	94.7 (p=0.3)*	94.7 (p=0.3)*	88.8 (p=0)

* is the method that gave the highest percentage of acceptance a null hypothesis for each situation.

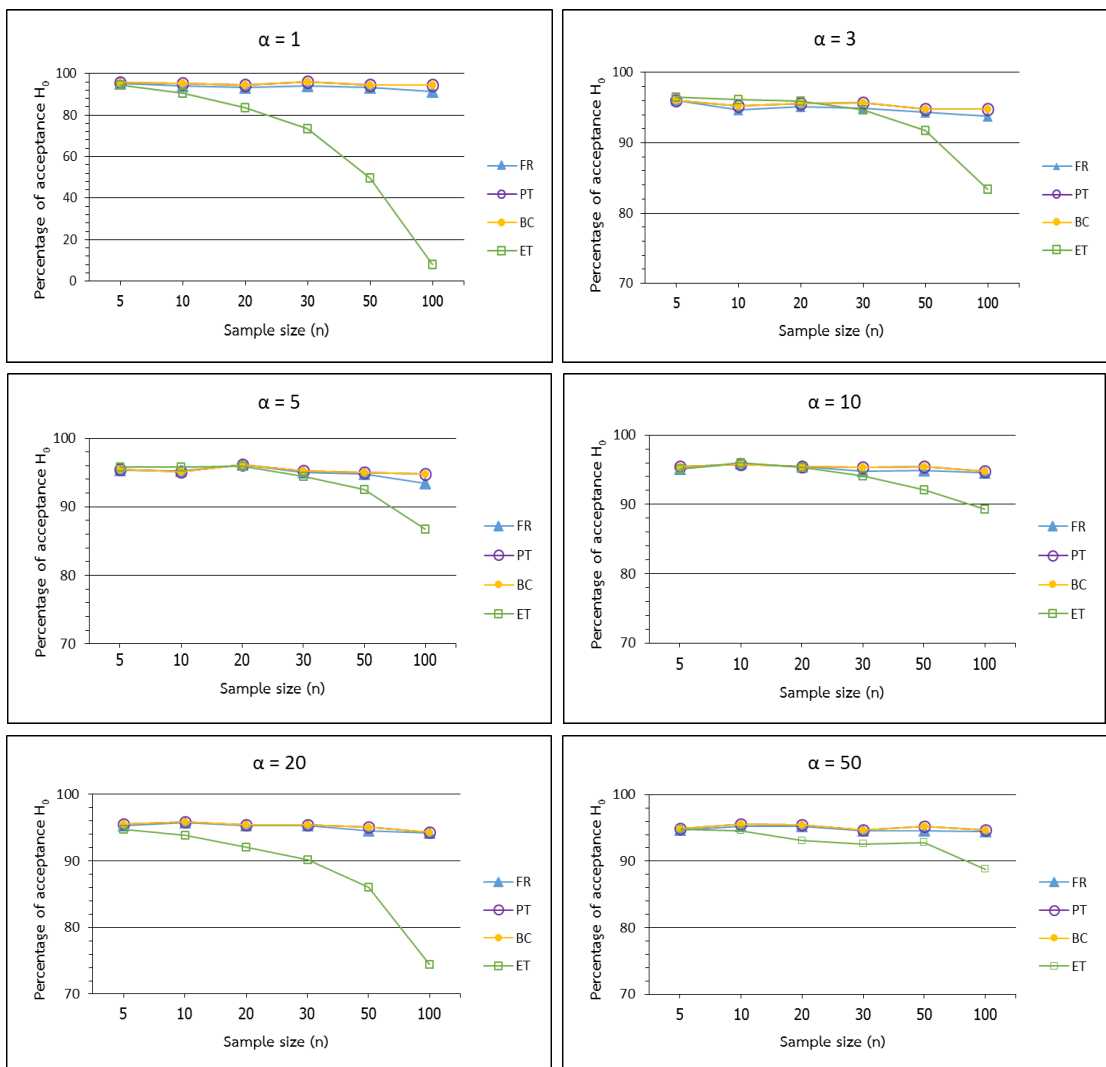


Figure 1 Percentages of acceptance a null hypothesis for Gamma distribution data with $\beta = 1$ and different values of α

เมื่อพิจารณาตารางที่ 1 คือ เมื่อข้อมูลมีการแจกแจงแกมมาด้วยพารามิเตอร์สเกล (β) เท่ากับ 1 และพารามิเตอร์รูปร่าง (α) แตกต่างกัน โดยพิจารณาจากค่าร้อยละการยอมรับสมมติฐานหลักพบว่าส่วนใหญ่วิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลัง มีแนวโน้มเป็นวิธีการแปลงที่มีประสิทธิภาพดีที่สุดเกือบทุกสถานการณ์ ยกเว้นกรณี α เท่ากับ 3 และ n เท่ากับ 5, 10, 20, กรณี α

เท่ากับ 5 และ n เท่ากับ 5, 10 และกรณี α เท่ากับ 10 และ n เท่ากับ 10 วิธีการแปลงแบบเลขชี้กำลังของ Manly จะเป็นวิธีการแปลงที่มีประสิทธิภาพดีที่สุด โดยค่า p ส่วนใหญ่ที่มีประสิทธิภาพดีสำหรับวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลัง เท่ากับ 0.3, 0.4, 0.5 และค่า p ส่วนใหญ่ที่มีประสิทธิภาพดีสำหรับวิธีการแปลงแบบเลขชี้กำลังของ Manly เท่ากับ -0.8, -0.2, -0.1 และ 0 และจากรูปที่ 1 จะเห็นได้ว่า

วิธีการแปลงแบบเลขชี้กำลังของ Manly มีแนวโน้มให้ค่าร้อยละการยอมรับสมมติฐานหลักลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้น แต่วิธีการแปลงแบบแบบ บ็อกซ์-ค็อกซ์ วิธีการแปลงแบบกำลัง และวิธีการแปลงแบบรากที่สองมีแนวโน้มให้ค่าร้อยละการยอมรับสมมติฐานหลักมีแนวโน้มคงเดิม แต่จะลดลงเล็กน้อยเมื่อ n เท่ากับ 50 และ 100

เมื่อพิจารณาดารางที่ 2 คือ เมื่อข้อมูลมีการแจกแจงแกมมาด้วยพารามิเตอร์สเกล (β) เท่ากับ 2

และพารามิเตอร์รูปร่าง (α) แตกต่างกัน โดยพิจารณาจากค่าร้อยละการยอมรับสมมติฐานหลักพบว่าส่วนใหญ่วิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลัง มีแนวโน้มเป็นวิธีการแปลงที่มีประสิทธิภาพที่สุดเกือบทุกสถานการณ์ ยกเว้นกรณี α เท่ากับ 3 และ n เท่ากับ 5, 10, 20, กรณี α เท่ากับ 5 และ n เท่ากับ 5, 10 วิธีการแปลงแบบเลขชี้กำลังของ Manly จะเป็นวิธีการแปลงที่มีประสิทธิภาพที่ดีที่สุด และกรณี α เท่ากับ 20 และ n เท่ากับ 20 วิธี

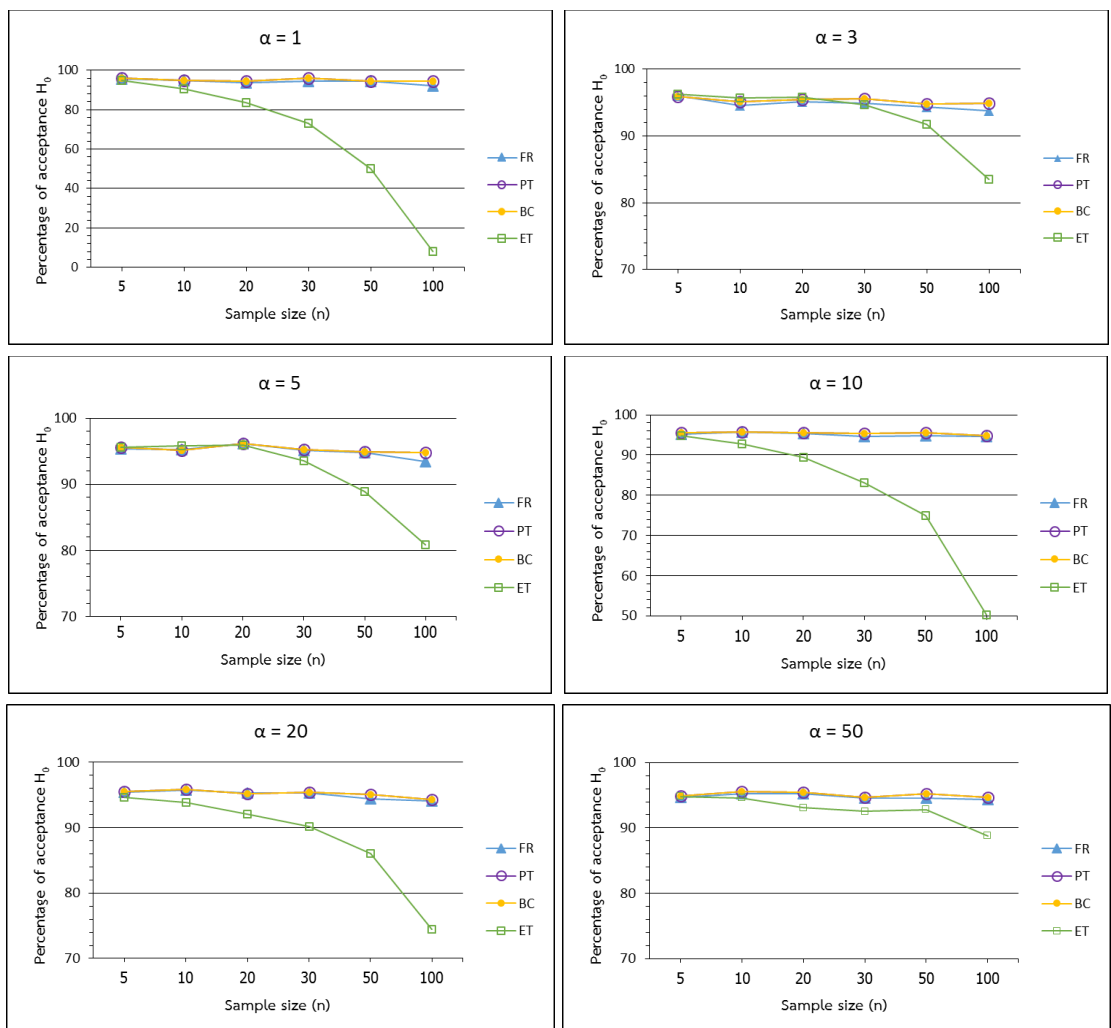


Figure 2 Percentages of acceptance a null hypothesis for Gamma distribution data with $\beta = 2$ and different values of α

Table 2 Percentages of acceptance a null hypothesis for Gamma distribution data with $\beta = 2$ and different values of α

α	Sample sizes (n)	Data transformation methods			
		FR method	PT method	BC method	ET method
1	5	95.7	95.9 (p=0.3)*	95.9 (p=0.3)*	94.7 (p=-0.4)
	10	94.9	95 (p=0.3)*	95 (p=0.3)*	90.6 (p=-0.4)
	20	93.6	94.4 (p=0.3)*	94.4 (p=0.3)*	83.5 (p=-0.4)
	30	94.6	95.9 (p=0.3)*	95.9 (p=0.3)*	73.1 (p=-0.4)
	50	94.3	94.5 (p=0.3)*	94.5 (p=0.3)*	50.2 (p=-0.4)
	100	92.1	94.3 (p=0.3)*	94.3 (p=0.3)*	8 (p=-0.4)
3	5	96	95.9 (p=0.5)	95.9 (p=0.5)	96.3 (p=-0.1)*
	10	94.6	95.1 (p=0.4)	95.1 (p=0.4)	95.7 (p=-0.1)*
	20	95.1	95.5 (p=0.3)	95.5 (p=0.3)	95.8 (p=-0.1)*
	30	94.9	95.6 (p=0.3, 0.4)*	95.6 (p=0.3, 0.4)*	94.7 (p=-0.1)
	50	94.3	94.8 (p=0.3)*	94.8 (p=0.3)*	91.7 (p=-0.1)
	100	93.8	94.9 (p=0.3)*	94.9 (p=0.3)*	83.5 (p=-0.1)
5	5	95.4	95.6 (p=0.4)*	95.6 (p=0.4)*	95.6 (p=-0.1)*
	10	95.2	95.1 (p=0.4, 0.5)	95.1 (p=0.4, 0.5)	95.8 (p=-0.1)*
	20	96.2*	96.2 (p=0.3)*	96.2 (p=0.3)*	95.9 (p=-0.1)
	30	95.1	95.2 (p=0.4)*	95.2 (p=0.4)*	93.5 (p=-0.1)
	50	94.8	94.9 (p=0.3)*	94.9 (p=0.3)*	88.9 (p=-0.1)
	100	93.4	94.8 (p=0.4)*	94.8 (p=0.4)*	80.8 (p=-0.1)
10	5	95.2	95.5 (p=0.7)*	95.5 (p=0.7)*	94.9 (p=0)
	10	95.8*	95.8 (p=0.2)*	95.8 (p=0.2)*	92.8 (p=0)
	20	95.4	95.5 (p=0.3)*	95.5 (p=0.3)*	89.5 (p=0)
	30	94.7	95.3 (p=0.5)*	95.3 (p=0.5)*	83.2 (p=0)
	50	94.9	95.5 (p=0.4)*	95.5 (p=0.4)*	75 (p=0)
	100	94.7	94.8 (p=0.3)*	94.8 (p=0.3)*	50.2 (p=0)
20	5	95.4	95.5 (p=0.4, 0.5)*	95.5 (p=0.4, 0.5)*	94.6 (p=0)
	10	95.7	95.9 (p=0.4, 0.5)*	95.9 (p=0.4, 0.5)*	93.9 (p=0)
	20	95.3*	95.2 (p=0.2)	95.2 (p=0.2)	92.1 (p=0)
	30	95.3	95.4 (p=0.3)*	95.4 (p=0.3)*	90.2 (p=0)
	50	94.4	95.1 (p=0.4)*	95.1 (p=0.4)*	86.1 (p=0)
	100	94.1	94.3 (p=0.3)*	94.3 (p=0.3)*	74.4 (p=0)
50	5	94.7	94.9 (p=0.6, 0.7, 0.9)*	94.9 (p=0.6, 0.7, 0.9)*	94.8 (p=0)
	10	95.2	95.6 (p=0.5, 0.6)*	95.6 (p=0.5, 0.6)*	94.6 (p=0)
	20	95.2	95.4 (p=0.3, 0.4)*	95.4 (p=0.3, 0.4)*	93.1 (p=0)
	30	94.6	94.7 (p=0.3)*	94.7 (p=0.3)*	92.5 (p=0)
	50	94.6	95.2 (p=0.6)*	95.2 (p=0.6)*	92.8 (p=0)
	100	94.3	94.7 (p=0.3)*	94.7 (p=0.3)*	88.8 (p=0)

* is the method that gave the highest percentage of acceptance a null hypothesis for each situation.

การแปลงแบบรากที่สี่จะเป็นวิธีการแปลงที่มีประสิทธิภาพที่สุด โดยค่า p ส่วนใหญ่ที่มีประสิทธิภาพดีสำหรับวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลัง คือ p เท่ากับ 0.3, 0.4 และ 0.5 และค่า p ส่วนใหญ่ที่มีประสิทธิภาพดีสำหรับวิธีการแปลงแบบเลขชี้กำลังของ Manly คือ p เท่ากับ -0.4, -0.1 และ 0 และจากรูปที่ 2 จะเห็นได้ว่าวิธีการแปลงแบบเลขชี้กำลังของ Manly มีแนวโน้มให้ค่าร้อยละการยอมรับสมมติฐานหลักลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น แต่วิธีการแปลงแบบแบบบ็อกซ์-ค็อกซ์ วิธีการแปลงแบบกำลัง และวิธีการแปลงแบบรากที่สี่มีแนวโน้มให้ค่าร้อยละการยอมรับสมมติฐานหลักมีแนวโน้มคงเดิม แต่จะลดลงเล็กน้อยเมื่อ n เท่ากับ 50 และ 100

เมื่อพิจารณาดารงที่ 1 และ 2 พบว่าค่าร้อยละการยอมรับสมมติฐานหลักที่ว่าข้อมูลที่แปลงแล้ว มีการแจกแจงปกติในทั้งสองตาราง ส่วนใหญ่ให้ผลลัพธ์ใกล้เคียงกันในเกือบทุกสถานการณ์ กล่าวคือ ผลลัพธ์เป็นไปในทิศทางเดียวกัน

4. สรุป

งานวิจัยนี้ใช้เกณฑ์การเปรียบเทียบประสิทธิภาพ คือ ค่าร้อยละการยอมรับสมมติฐานหลักที่ว่าข้อมูลที่แปลงแล้วมีการแจกแจงปกติ โดยใช้สถิติทดสอบแอนเดอร์สัน-ดาร์ลิง ที่ระดับนัยสำคัญ 0.05 นั่นคือ วิธีการแปลงที่มีประสิทธิภาพจะพิจารณาจากข้อมูลที่ศึกษาซึ่งเดิมมีการแจกแจงแกมมา พบว่าวิธีการแปลงข้อมูลที่ทำให้ข้อมูลมีการแจกแจงปกติหรือเข้าใกล้การแจกแจงปกติมากที่สุดในแต่ละสถานการณ์สามารถสรุปได้ดังตารางที่ 3 และ 4

กรณีที่ n มีขนาดใหญ่ ($n \geq 30$) จากทั้งตารางที่ 3 และ 4 จะพบว่าวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลังมีประสิทธิภาพดีในทุกระดับของ α และ β โดยส่วนใหญ่ p ที่เหมาะสมเท่ากับ 0.3

และในกรณีที่ n มีขนาดเล็ก ($n < 30$) จากตารางที่ 3 และ 4 จะพบว่าที่ระดับ $\alpha = 1, 10, 20$ และ 50 วิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลังส่วนใหญ่จะยังคงมีประสิทธิภาพดีในทุกระดับของ β ยกเว้นที่ $\alpha = 3, 5$ พบว่าวิธีการแปลงแบบเลขชี้กำลังของ Manly จะมีประสิทธิภาพดีกว่า โดย p ที่เหมาะสมสำหรับวิธีการแปลงแบบเลขชี้กำลังของ Manly ส่วนใหญ่เท่ากับ -0.1

เมื่อพิจารณาการแจกแจงแกมมาด้วยพารามิเตอร์สเกล (β) เท่ากับ 1 และ 2 ทุกระดับพารามิเตอร์รูปร่าง (α) พบว่า β ที่แตกต่างกันไม่มีผลต่อประสิทธิภาพของวิธีการแปลงข้อมูลแต่ละวิธีที่ศึกษาในครั้งนี้ เนื่องจากให้ผลลัพธ์ค่อนข้างคล้ายกัน

5. วิจารณ์

ผลการเปรียบเทียบประสิทธิภาพวิธีการแปลงข้อมูลทั้ง 4 วิธี พบว่าวิธีการแปลงแบบบ็อกซ์-ค็อกซ์และวิธีการแปลงแบบกำลังมีแนวโน้มให้ร้อยละของการแปลงข้อมูลได้สำเร็จเกือบทุกสถานการณ์ เมื่อเลขกำลังของการแปลงข้อมูล (p) ส่วนใหญ่มีค่าเท่ากับ 0.3, 0.4 ซึ่งสอดคล้องกับงานวิจัยของ Paiboolworachat [4] และ Chortirat [5] ส่วนวิธีการแปลงแบบรากที่สี่ เมื่อ n เพิ่มขึ้นจะให้ร้อยละของการแปลงข้อมูลได้สำเร็จลดลง ซึ่งสอดคล้องกับงานวิจัยของ Kaewprasert และคณะ [6] นอกจากนี้วิธีการแปลงแบบกำลังให้ร้อยละของการแปลงข้อมูลได้สำเร็จมาก เมื่อ n เพิ่มขึ้น และ p เท่ากับ 0.3, 0.4, 0.5, 0.6 ซึ่งสอดคล้องกับงานวิจัยของ Krataithong [1] และวิธีการแปลงแบบเลขชี้กำลังของ Manly จะให้ร้อยละของการแปลงข้อมูลได้สำเร็จมาก เมื่อ p เท่ากับ -0.1, -0.2, -0.3 แต่เมื่อ n มีขนาดเพิ่มขึ้นจะให้ร้อยละของการแปลงข้อมูลได้สำเร็จลดลง ซึ่งสอดคล้องกับงานวิจัยของ Kerdsawang [3]

Table 3 The best data transformation method for gamma distribution data with $\beta = 1$ and different values of α

Shape parameters (α)	Sample sizes (n)					
	5	10	20	30	50	100
1	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)
3	ET (-0.3)	ET (-0.3)	ET (-0.2)	BC, PT (0.4)	BC, PT (0.3)	BC, PT (0.3)
5	ET (-0.1)	ET (-0.2)	BC, PT (0.3)	BC, PT (0.4)	BC, PT (0.3)	BC, PT (0.4)
10	BC, PT (0.7)	ET (-0.1)	BC, PT (0.3)	BC, PT (0.4)	BC, PT (0.4)	BC, PT (0.3)
20	BC, PT (0.4, 0.5)	BC, PT (0.4, 0.5)	BC, PT (0.2)	BC, PT (0.3)	BC, PT (0.4)	BC, PT (0.3)
50	BC, PT (0.7, 0.9)	BC, PT (0.5, 0.6)	BC, PT (0.4)	BC, PT (0.3)	BC, PT (0.6)	BC, PT (0.3)

The values in parenthesis are the suitable power of transformation (p) for each method.

Table 4 The best data transformation method for Gamma distribution data with $\beta = 2$ and different values of α

Shape parameters (α)	Sample sizes (n)					
	5	10	20	30	50	100
1	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)	BC, PT (0.3)
3	ET (-0.1)	ET (-0.1)	ET (-0.1)	BC, PT (0.3, 0.4)	BC, PT (0.3)	BC, PT (0.3)
5	ET (-0.1) BC, PT (0.4)	ET (-0.1)	FR, BC, PT (0.3)	BC, PT (0.4)	BC, PT (0.3)	BC, PT (0.4)
10	BC, PT (0.7)	FR, BC, PT (0.2)	BC, PT (0.3)	BC, PT (0.5)	BC, PT (0.4)	BC, PT (0.3)
20	BC, PT (0.4, 0.5)	BC, PT (0.4, 0.5)	FR	BC, PT (0.3)	BC, PT (0.4)	BC, PT (0.3)
50	BC, PT (0.6, 0.7, 0.9)	BC, PT (0.5, 0.6)	BC, PT (0.3, 0.4)	BC, PT (0.3)	BC, PT (0.6)	BC, PT (0.3)

The values in parenthesis are the suitable power of transformation (p) for each method.

6. ข้อเสนอแนะ

ผู้ที่สนใจศึกษาการแก้ปัญหาข้อมูลที่ไม่มี การแจกแจงปรกติด้วยวิธีการแปลงข้อมูลอาจศึกษาวิธีการแปลงแบบอื่น ๆ เพื่อเปรียบเทียบประสิทธิภาพกับวิธีการแปลงแบบกำลังและแบบบ็อกซ์-ค็อกซ์ที่ให้ ร้อยละการยอมรับที่สูง เช่น วิธีการแปลงแบบ Aranda-Ordaz [12] หรือผู้ที่สนใจอาจศึกษากรณีที่ ข้อมูลมีการแจกแจงแบบอื่น ๆ ได้แก่ การแจกแจงแบบ

ลอจิสติก (logistic distribution) [13] การแจกแจงลิ อกนอร์มอล (Lognormal distribution) [13] เป็นต้น นอกจากนี้วิธีการแปลงข้อมูลทั้ง 4 วิธี ที่ได้ศึกษาครั้งนี้ พบว่าวิธีการแปลงแบบรากที่มีแนวโน้มให้ผลไม่ค่อยดี นัก แต่วิธีคำนวณไม่ซับซ้อน ส่วนวิธีการแปลงแบบ บ็อกซ์-ค็อกซ์และวิธีการแปลงแบบเลขชี้กำลังของ Manly มีการคำนวณที่ซับซ้อนกว่าแต่ให้ผลที่ค่อนข้าง ดี โดยวิธีการแปลงแบบเลขชี้กำลังของ Manly ให้ผลที่

ดีในตัวอย่างขนาดเล็ก ส่วนวิธีการแปลงแบบกำลังนั้นมีวิธีคำนวณที่ไม่ซับซ้อนอีกทั้งยังให้ผลที่ดี ในการพิจารณาการเลือกใช้วิธีการแปลงข้อมูลนั้น ผู้สนใจศึกษาควรเลือกใช้วิธีการที่ไม่ซับซ้อน ตรงกับลักษณะของข้อมูลที่ต้องการและให้ผลที่ดี เช่น กรณีที่ข้อมูลมีขนาดเล็กควรใช้วิธีการแปลงแบบเลขชี้กำลังของ Manly หรือกรณีที่ข้อมูลมีขนาดใหญ่ควรใช้วิธีการแปลงแบบบ็อกซ์-ค็อกซ์จึงจะได้ผลดีที่สุด

7. References

- [1] Krataithong, N., 1999, Data Transformation to Normal Distribution, Master Thesis, Chulalongkorn University, Bangkok, 303 p. (in Thai)
- [2] Ngamjarus, C., 2001, The Comparison of Transformation Methods for Exponential Data to Normal Data, Master Thesis, King Mongkut's University of Technology North Bangkok, Bangkok, 190 p. (in Thai)
- [3] Kerdawang, S., 2005, Comparison of Data Transformation Techniques for Normality, Master Thesis, Chulalongkorn University, Bangkok, 238 p. (in Thai)
- [4] Paiboolworachat, J., 2011, Data the Efficiency Comparisons Among the Box-Cox Power Transformation Method and the New Power Transformation Method for Exponential Distribution, Special Problem, Burapha University, Chonburi, 40 p. (in Thai)
- [5] Chortirat, T., 2011, A Comparison of The Four Data Transformation Methods for Weibull Districtbuted Data, Master Thesis, Kasetsart University, Bangkok, 235 p. (in Thai)
- [6] Kaewprasert, T., Khamkong, M. and Book kamana, P., 2017, A comparison of data transformation methods of generalized exponential distribution and estimation of summer rainfall in Chiang Dao, Chiang Mai, Burapha Sci. J. 22(3): 385-396.
- [7] Channarong, C., 2005, Comparison of Power of the Test for Four Normality Testing, Master Thesis, Kasetsart University, Bangkok, 224 p. (in Thai)
- [8] Pinthongpan, A., 2014, Comparison of The Goodness of Fit Tests for Normality, Master Thesis, Kasetsart University, Bangkok, 151 p. (in Thai)
- [9] Montgomery, D.C., 2012, Introduction to Statistical Quality Control, 7th Ed., John Wiley & Sons, New York, 774 p.
- [10] Anderson, T.W. and Darling, D.A., 1952, Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes, Ann. Math. Stat. 23: 193-212.
- [11] Shorak, G.R. and Wellner, J.A., 1986, Empirical Processes with Applications to Statistics, Wiley, New York, 239 p.
- [12] Aranda-Ordaz, F.J., 1981, On two families of transformations to additivity for binary response data, Biometrika. 68: 357-363.
- [13] Forbes, C., Evans, M., Hastings, N. and Peacock, B., 2011, Statistical Distributions, 4th Ed., John Wiley & Sons, New Jersey, 212 p.