

การเปรียบเทียบประสิทธิภาพของการแจกแจงปัวซอง-ลินด์เลย์
โดยใช้ชุดข้อมูลจริง
Comparative Performance of Poisson-Lindley Distributions
Using Real Data Sets

กมลรัตน์ ไทรทอง และ ชฎารัตน์ ถาปัน*

Kamonrat Thaithong and Chadarat Tapan*

สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม จ.พิษณุโลก ประเทศไทย
Department of Mathematics, Faculty of Science and Technology, Pibulsongkram Rajabhat University,
Phitsanulok, Thailand

วันที่ส่งบทความ : 16 กุมภาพันธ์ 2569 วันที่แก้ไขบทความ : 8 มิถุนายน 2569 วันที่ตอบรับบทความ : 9 มิถุนายน 2569

Received: 16 February 2026, Revised: 8 June 2026, Accepted: 9 June 2026

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการแจกแจงปัวซอง-ลินด์เลย์ 3 รูปแบบ ได้แก่ การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ โดยเปรียบเทียบเชิงประจักษ์กับการแจกแจงปัวซองและทวินามลบซึ่งเป็นแบบจำลองพื้นฐานสำหรับข้อมูลเชิงนับ การศึกษานี้ใช้ชุดข้อมูลจริงจำนวน 3 ชุด ได้แก่ จำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนของจังหวัดระยอง จำนวนผู้เสียชีวิตจากโรคติดเชื้อไวรัสโคโรนา 2019 ของจังหวัดศรีสะเกษ จากเว็บไซต์ศูนย์กลางข้อมูลเปิดภาครัฐ และจำนวนครั้งในการพบแพทย์จากการสำรวจสุขภาพแห่งชาติของประเทศออสเตรเลียจากแฟ้มเก็บ AER ในโปรแกรม R การประมาณค่าพารามิเตอร์ดำเนินการด้วยวิธีภาวะน่าจะเป็นสูงสุด และประเมินความเหมาะสมของการแจกแจงโดยใช้ค่า -Log-likelihood (-LL) ค่า Akaike's Information Criterion (AIC) ค่า Bayesian Information Criterion (BIC) และสถิติทดสอบ Kolmogorov-Smirnov (KS) ผลการศึกษาพบว่า การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปแสดงถึงศักยภาพในการอธิบายข้อมูลเชิงนับทั้ง 3 ชุด ที่มีลักษณะการกระจายเกินเกณฑ์ได้เป็นอย่างดี

คำสำคัญ : การแจกแจงปัวซองผสม การแจกแจงลินด์เลย์ การกระจายเกินเกณฑ์ การประมาณค่าความน่าจะเป็นสูงสุด

Abstract

This study aims to compare the performance of three distributions of the Poisson-Lindley distribution, namely the generalized Poisson-Lindley distribution, the weighted

*ที่อยู่ติดต่อ E-mail address: chadarat@psru.ac.th

<https://doi.org/10.55003/scikmitl.2026.270667>

Poisson-Lindley distribution, and the three-parameter Poisson-Lindley distribution. The comparative analysis is conducted against the Poisson and negative binomial distributions, which are commonly used as baseline models for count data. The study utilizes three real data sets: the number of road traffic accident fatalities in Rayong Province, the number of deaths from coronavirus disease 2019 in Sisaket province obtained from the open government data center of Thailand, and the number of physician visits from the Australian national health survey available in the AER package in the R program. Parameter estimation is performed using the maximum likelihood method. The goodness-of-fit of the models is evaluated using the -Log-likelihood value (-LL), Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), and the Kolmogorov-Smirnov (KS) test statistic. The results show that the generalized Poisson-Lindley distribution provides the most suitable fit for three count data sets exhibiting overdispersion.

Keywords: Mixed Poisson distribution, Lindley distribution, Overdispersion, Maximum likelihood estimation

1. บทนำ

ข้อมูลเชิงนับ (Count data) คือข้อมูลที่แสดงจำนวนครั้งของการเกิดเหตุการณ์หนึ่ง ๆ ซึ่งพบได้ในหลายสาขา เช่น การประกันภัย การแพทย์ สาธารณสุข เกษตรกรรม ชีววิทยา และวิศวกรรมศาสตร์ การแจกแจงที่นิยมใช้ในการวิเคราะห์ข้อมูลประเภทนี้คือ การแจกแจงปัวซอง (Poisson distribution) ซึ่งตั้งอยู่บนสมมติฐานว่าค่าเฉลี่ยและค่าความแปรปรวนมีค่าเท่ากัน (Equidispersion) อย่างไรก็ตามในทางปฏิบัติมักพบว่า ข้อมูลเชิงนับมีลักษณะการกระจายเกินเกณฑ์ (Overdispersion) กล่าวคือ ค่าความแปรปรวนสูงกว่าค่าเฉลี่ย หรือในบางกรณีอาจพบการกระจายต่ำกว่าเกณฑ์ (Underdispersion) ซึ่งมีค่าความแปรปรวนต่ำกว่าค่าเฉลี่ย ส่งผลให้การใช้การแจกแจงปัวซองไม่เหมาะสมในการวิเคราะห์ข้อมูลดังกล่าว โดยทั่วไปการแจกแจงทวินามลบ (Negative Binomial: NB) มักถูกใช้เป็นการแจกแจงมาตรฐานสำหรับข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ เนื่องจากสามารถอธิบายความแปรปรวนที่มากกว่าค่าเฉลี่ยได้ และเพื่อแก้ไขข้อจำกัดนี้จึงมีการพัฒนาการแจกแจงปัวซองผสม (Mixed Poisson distributions) ซึ่งสามารถรองรับข้อมูลที่มีค่าความแปรปรวนมากหรือน้อยกว่าค่าเฉลี่ยได้อย่างยืดหยุ่น และเหมาะสมสำหรับข้อมูลเชิงนับที่มีค่าเป็นจำนวนเต็มไม่เป็นลบ ซึ่งโดยทั่วไปมักไม่เป็นไปตามการแจกแจงปกติ (Tharshan & Wijekoon, 2022b)

ความสำคัญของการพัฒนาการแจกแจงปัวซองผสมดังกล่าวได้นำไปสู่การศึกษาวิจัยจำนวนมากในช่วงที่ผ่านมา โดยเฉพาะงานวิจัยที่มุ่งเน้นการสร้างแบบจำลองเพื่อจัดการกับปัญหาการกระจายเกินเกณฑ์ของข้อมูลเชิงนับ ซึ่งไม่สามารถอธิบายได้ด้วยการแจกแจงปัวซอง อันเนื่องมาจากสมมติฐานที่กำหนดให้ค่าเฉลี่ยและค่าความแปรปรวนต้องมีค่าเท่ากัน แนวคิดของการแจกแจงปัวซองผสม คือ การกำหนดให้พารามิเตอร์ของการแจกแจงปัวซองเป็นตัวแปรสุ่มที่ขึ้นอยู่กับการแจกแจงอื่น ดังนั้น การเลือกการแจกแจงที่ใช้เป็นตัวผสมจึงมีบทบาทสำคัญต่อความยืดหยุ่นและประสิทธิภาพของแบบจำลอง (Cameron & Trivedi, 2014)

งานวิจัยในช่วงทศวรรษที่ผ่านมาได้เสนอการแจกแจงปัวซองผสมหลากหลายรูปแบบ ตัวอย่างเช่น การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป (Generalized Poisson-Lindley distribution: GPL) ที่เสนอโดย Bhati et al. (2015) เพื่อเพิ่มความยืดหยุ่นในการจำลองข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์และเบ้ขวา โดยทำการประมาณค่าพารามิเตอร์ด้วยวิธีโมเมนต์และอัลกอริทึม EM ต่อมา Das et al. (2018) และคณะได้เสนอการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ (Three-Parameter Poisson-Lindley distribution: TPPLD) สำหรับข้อมูลที่มีการกระจายเกินเกณฑ์ และประเมินพารามิเตอร์ด้วยวิธีโมเมนต์และวิธีความน่าจะเป็นสูงสุด ผลการประยุกต์ใช้กับข้อมูลจริง พบว่า การแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์มีค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ต่ำที่สุดเมื่อเทียบกับการแจกแจงอื่น นอกจากนี้ Atikankul et al. (2020) ได้เสนอการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก (Poisson-Weighted Lindley distribution: PWL) โดยประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด และประเมินประสิทธิภาพของแบบจำลองด้วยเกณฑ์ค่า Akaike's Information Criterion (AIC) ค่า Bayesian Information Criterion (BIC) และเกณฑ์สารสนเทศของอะกะอิเกะแบบแก้ไขแล้ว (Corrected Akaike Information Criterion: AICc) ผลการศึกษาชี้ให้เห็นว่า การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนักมีความเหมาะสมอย่างยิ่งสำหรับการวิเคราะห์ข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์

แนวโน้มการพัฒนาการแจกแจงปัวซองผสมยังคงดำเนินต่อเนื่องในช่วงไม่กี่ปีที่ผ่านมา โดยมีมีการนำเสนอการแจกแจงแบบใหม่ที่มีความยืดหยุ่นสูงขึ้นและสามารถรองรับรูปแบบการกระจายของข้อมูลเชิงนับได้หลากหลายยิ่งขึ้น โดย Ahsan-ul-Haq et al. (2022) ได้เสนอการแจกแจงปัวซอง-เอ็กซ์แอล (Poisson-XL distribution) ซึ่งเกิดจากการผสมการแจกแจงปัวซองกับการแจกแจงเอ็กซ์แอลลินด์เลย์ (XLindley distribution) เพื่อข้อมูลเชิงนับที่มีลักษณะการเบ้ขวาโดยเฉพาะข้อมูลที่มีการกระจายเท่ากันหรือการกระจายเกินเกณฑ์ ผลการศึกษาชี้ให้เห็นถึงความยืดหยุ่นและประสิทธิภาพของการแจกแจงดังกล่าวในการประยุกต์ใช้กับข้อมูลจริง นอกจากนี้ Tharshan & Wijekoon (2022a) ได้นำเสนอการแจกแจงปัวซอง-ลินด์เลย์แบบดัดแปลง (Poisson-Modified Lindley distribution) ซึ่งเป็นการแจกแจงไม่ต่อเนื่องที่ออกแบบมาเพื่อการวิเคราะห์ข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์ โดยได้ศึกษาเชิงลึกทั้งคุณสมบัติทางสถิติและการประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุดและวิธีโมเมนต์ ผลการวิเคราะห์ พบว่า การแจกแจงดังกล่าวให้ประสิทธิภาพที่เหนือกว่าการแจกแจงปัวซองและการแจกแจงปัวซอง-ลินด์เลย์แบบดั้งเดิม จึงเป็นอีกทางเลือกหนึ่งที่เหมาะสมสำหรับการวิเคราะห์ข้อมูลเชิงนับประเภทดังกล่าว

ต่อมา Atikankul (2023) ได้พัฒนาการแจกแจงปัวซอง-ลินด์เลย์ทั่วไป (Poisson-generalized Lindley distribution: PGL) สำหรับข้อมูลที่มีการกระจายเกินเกณฑ์ โดยศึกษาคุณสมบัติทางทฤษฎีและประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด ผลการประยุกต์ใช้กับข้อมูลจริงชี้ให้เห็นว่า การแจกแจงดังกล่าวให้ค่า AIC และ BIC ต่ำที่สุด รวมทั้งมีค่า LL (Log-likelihood) และค่าพี (p-value) จากการทดสอบ Anderson-Darling (AD) สูงกว่าการแจกแจงปัวซองและทวินามลบ แสดงให้เห็นถึงความเหมาะสมของการแจกแจงในการใช้งานจริง นอกจากนี้ ในปี Alomair et al. (2024) ได้เสนอการแจกแจงปัวซอง-ลินด์เลย์ดีกรีสองแบบปรับปรุง (Poisson-Improved Second-Degree Lindley) เพื่อแก้ไขปัญหาการกระจายเกินเกณฑ์ของข้อมูลเชิงนับ โดยประมาณค่าพารามิเตอร์ด้วยวิธีโมเมนต์และวิธีภาวะน่าจะเป็นสูงสุด ผลการศึกษาทั้งจากการทดลองเชิงจำลองและข้อมูลทางการแพทย์พบว่า การแจกแจงที่นำเสนอมีคุณสมบัติไม่

เอนเอียงและให้การประมาณค่าที่มีความน่าเชื่อถือ ส่งเสริมให้การแจกแจงนี้เป็นอีกหนึ่งทางเลือกที่มีศักยภาพในการวิเคราะห์ข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์

จากการทบทวนวรรณกรรม พบว่า งานวิจัยที่เกี่ยวข้องส่วนใหญ่ให้ความสำคัญกับการพัฒนาแบบจำลองการแจกแจงปัวซอง-ลินด์เลย์ในรูปแบบใหม่ และการศึกษาคุณสมบัติทางทฤษฎีของการแจกแจงดังกล่าว เช่น การศึกษาฟังก์ชันมวลความน่าจะเป็น ค่าเฉลี่ย ความแปรปรวน และวิธีการประมาณค่าพารามิเตอร์ อย่างไรก็ตาม งานวิจัยจำนวนมากมักพิจารณาประสิทธิภาพของแบบจำลองเพียงรูปแบบเดียวหรือเปรียบเทียบกับแบบจำลองพื้นฐานในการประเมินความเหมาะสมของแบบจำลอง ส่งผลให้ยังขาดการศึกษาเชิงเปรียบเทียบระหว่างแบบจำลองในกลุ่มการแจกแจงปัวซอง-ลินด์เลย์ที่ได้รับการพัฒนาขึ้นในช่วงเวลาต่าง ๆ โดยเฉพาะเมื่อใช้กับชุดข้อมูลจริง ดังนั้น งานวิจัยนี้จึงมุ่งเติมเต็มช่องว่างขององค์ความรู้ดังกล่าว โดยทำการเปรียบเทียบประสิทธิภาพของการแจกแจงปัวซอง-ลินด์เลย์ 3 รูปแบบ ได้แก่ การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป (Bhati et al., 2015) การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก (Atikankul et al., 2020) และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ (Das et al., 2018) เนื่องจากการแจกแจงเหล่านี้เป็นการขยายรูปแบบของการแจกแจงปัวซอง-ลินด์เลย์พื้นฐานที่ได้รับการนำเสนอในงานวิจัยก่อนหน้า และมีความยืดหยุ่นในการอธิบายข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ การเลือกการแจกแจงทั้งสามรูปแบบนี้จึงช่วยให้สามารถเปรียบเทียบประสิทธิภาพของการขยายการแจกแจงในตระกูลเดียวกันได้อย่างเป็นระบบ โดยเปรียบเทียบกับ การแจกแจงปัวซองและการแจกแจงทวินามลบด้วยชุดข้อมูลจริงจำนวน 3 ชุด ในการประเมินประสิทธิภาพของการแจกแจงได้ใช้เกณฑ์เปรียบเทียบ ได้แก่ ค่า -LL AIC BIC และการทดสอบคอลโมโกรอฟ-สมิรโนฟ (Kolmogorov-Smirnov: KS) เพื่อประเมินความเหมาะสมและความสามารถของแบบจำลองในการจำลองข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์

2. วิธีการดำเนินการวิจัย

ในการศึกษาครั้งนี้ได้เลือกพิจารณาการแจกแจงในกลุ่มปัวซอง-ลินด์เลย์จำนวน 3 รูปแบบ ได้แก่ การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ เนื่องจากการแจกแจงดังกล่าวได้รับการพัฒนาในช่วงเวลาที่แตกต่างกัน และมีโครงสร้างพารามิเตอร์ที่สะท้อนแนวทางการขยายการแจกแจงปัวซอง-ลินด์เลย์ในลักษณะที่แตกต่างกัน กล่าวคือ การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปเป็นการขยายการแจกแจงพื้นฐานเพื่อเพิ่มความยืดหยุ่นของรูปแบบ ส่วนการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนักเป็นการปรับโครงสร้างของการแจกแจงผ่านแนวคิดการถ่วงน้ำหนักเพื่อรองรับข้อมูลที่มีการกระจายเกินเกณฑ์ ในขณะที่ การแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์เพิ่มจำนวนพารามิเตอร์เพื่อเพิ่มความสามารถในการปรับรูปร่างของการแจกแจงให้สอดคล้องกับข้อมูลจริง ดังนั้น การเลือกการแจกแจงทั้งสามรูปแบบจึงช่วยให้สามารถเปรียบเทียบแนวทางการพัฒนาการแจกแจงปัวซอง-ลินด์เลย์ที่สำคัญ ภายใต้กรอบการวิเคราะห์เดียวกัน ทั้งในมิติของการเพิ่มพารามิเตอร์และการปรับโครงสร้างของการแจกแจงจึงมีความเหมาะสมสำหรับการนำมาใช้ในการศึกษาเชิงเปรียบเทียบประสิทธิภาพของการแจกแจงในงานวิจัยนี้

2.1 การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป

การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปเป็นการผสมผสานระหว่างการแจกแจงปัวซองและการแจกแจงลินด์เลย์ เพื่อเพิ่มความยืดหยุ่นในการอธิบายข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ การแจกแจงดังกล่าวถูกกำหนดด้วยพารามิเตอร์สองตัว ซึ่งช่วยให้สามารถปรับปรุงร่างของการแจกแจงให้สอดคล้องกับโครงสร้างของข้อมูลได้อย่างเหมาะสม และเพิ่มประสิทธิภาพในการจำลองข้อมูล (Bhati et al., 2015)

ทฤษฎีบท 2.1 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปัวซอง-ลินด์เลย์แบบทั่วไป และมี θ และ α เป็นพารามิเตอร์ แล้วฟังก์ชันมวลความน่าจะเป็น (Probability Mass Function: PMF) ของ X แสดงดังสมการที่ (1)

$$f(x) = \frac{\theta^2}{(\theta+\alpha)(1+\theta)^{x+1}} \left(1 + \frac{\alpha(x+1)}{1+\theta}\right), x = 0, 1, 2, \dots \quad (1)$$

โดยที่ $\theta > 0$ และ $\alpha > 0$

ค่าเฉลี่ยและความแปรปรวนสำหรับตัวแปรสุ่ม X แสดงดังสมการที่ (2) และ (3) ตามลำดับ

$$E(X) = \frac{2\alpha+\theta}{\theta(\alpha+\theta)} \quad (2)$$

$$Var(X) = \frac{2\alpha^2(1+\theta)+\theta^2(1+\theta)+\alpha\theta(4+3\theta)}{\theta^2(\alpha+\theta)^2} \quad (3)$$

ดัชนีการกระจายตัว (Index of Dispersion: ID) แสดงดังสมการที่ (4)

$$ID = \frac{2\alpha^2(\theta+1)+\alpha\theta(3\theta+4)+\theta^2(\theta+1)}{\theta(\alpha+\theta)(2\alpha+\theta)} \quad (4)$$

การประมาณค่าพารามิเตอร์ของตัวอย่างสุ่ม $X_1, X_2, \dots, X_n$ ที่ได้จากการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปดำเนินการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation: MLE) ซึ่งมีฟังก์ชันล็อกภาวะน่าจะเป็น แสดงดังสมการที่ (5)

$$\begin{aligned} \ln L(x; \alpha, \theta) &= 2n \ln \theta - n \ln(\theta + \alpha) - \sum_{i=1}^n (x_i + 1) \ln(1 + \theta) \\ &\quad + \sum_{i=1}^n \left(1 + \frac{\alpha(x_i+1)}{1+\theta}\right) \end{aligned} \quad (5)$$

2.2 การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก

การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนักเป็นแบบจำลองที่พัฒนาขึ้นเพื่อเพิ่มความยืดหยุ่นในการวิเคราะห์ข้อมูลเชิงนับ โดยเกิดจากการผสมผสานระหว่างการแจกแจงปัวซองและการแจกแจงลินด์เลย์แบบ

ถ่วงน้ำหนัก เพื่อให้การแจกแจงสามารถรองรับระดับความแปรปรวนที่สูงกว่ามาตรฐานได้อย่างเหมาะสม ความยืดหยุ่นที่เพิ่มขึ้นดังกล่าวช่วยเสริมศักยภาพของการแจกแจงในการอธิบายข้อมูลที่มีลักษณะการกระจายเกินเกณฑ์ได้ดีกว่าแบบจำลองดั้งเดิม (Atikankul et al., 2020)

ทฤษฎีบท 2.2 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และมี θ และ α เป็นพารามิเตอร์แล้วฟังก์ชันมวลความน่าจะเป็น ของ X แสดงดังสมการที่ (6)

$$f(x) = \frac{\theta^2(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)} \left(\frac{\theta+x+2}{(\theta+1)^{x+2}} - \frac{\alpha\theta+\theta+x+2}{(\alpha\theta+\theta+1)^{x+2}} \right), x = 0, 1, 2, \dots \quad (6)$$

โดยที่ $\theta > 0$ และ $\alpha > 0$

ค่าเฉลี่ยและความแปรปรวนสำหรับตัวแปรสุ่ม แสดงดังสมการที่ (7) และ (8) ตามลำดับ

$$E(X) = \delta_1 \delta_3 \quad (7)$$

$$Var(X) = \delta_1 \left(\frac{\theta^2+4\theta+6}{\theta^2} - \frac{\delta_2^2+4\delta_2+6}{\delta_2^2} - \delta_1 \delta_3^2 \right) \quad (8)$$

$$\text{เมื่อ } \delta_1 = \frac{\theta^2+(1+\alpha)^2}{\alpha\theta(1+\alpha)+\alpha(2+\alpha)}, \delta_2 = \alpha\theta + \theta \text{ และ } \delta_3 = \frac{\theta+2}{\theta^3} - \frac{\alpha\theta+\theta+2}{(\alpha\theta+\theta)^3}$$

ดัชนีการกระจายตัว แสดงดังสมการที่ (9)

$$ID = \frac{-\theta^4\delta_3^2\delta_1\delta_2^4 - \theta^4\delta_2^2 - 6\theta^4\delta_2 - 6\theta^4 + \theta^2\delta_2^4 - 4\theta\delta_2^4 + 6\delta_2^4}{\theta^4\delta_3\delta_2^4} \quad (9)$$

การประมาณค่าพารามิเตอร์ของตัวอย่างสุ่ม $X_1, X_2, \dots, X_n$ ที่ได้จากการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนักดำเนินการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด ซึ่งมีฟังก์ชันลือคภาวน่าจะเป็น แสดงดังสมการที่ (10)

$$\begin{aligned} \ln L(x; \alpha, \theta) = & 2n \ln \alpha + \sum_{i=1}^n \ln(x_i + \alpha^2 + 3(1 + \alpha)) - \sum_{i=1}^n (4 + x_i) \ln(1 + \alpha) \\ & - (\sum_{i=1}^n x_i + 2n) \ln(\theta + 1) - (\sum_{i=1}^n x_i + 2n) \ln(\alpha\theta + 1) \\ & + \sum_{i=1}^n \ln((\theta + x_i + 2)(\alpha\theta + \theta + 1)^{x_i+2} - (\alpha\theta + \theta + x_i + 2)^{x_i+2}) \end{aligned} \quad (10)$$

2.3 การแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์

การแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์เป็นการขยายการแจกแจงเพื่อเพิ่มความยืดหยุ่นในการอธิบายข้อมูลเชิงนับ โดยเฉพาะในกรณีที่ข้อมูลมีลักษณะการกระจายเกินเกณฑ์หรือการกระจายต่ำกว่า

เกณฑ์ การเพิ่มพารามิเตอร์ตัวที่สามช่วยให้การแจกแจงสามารถปรับปรุงให้สอดคล้องกับโครงสร้างของข้อมูลจริงได้ดียิ่งขึ้น ส่งผลให้การแจกแจงนี้มีศักยภาพในการรองรับระดับความแปรปรวนของข้อมูลที่หลากหลายมากกว่าการแจกแจงดั้งเดิม (Das et al., 2018)

ทฤษฎีบท 3. ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ และมี θ , β และ α เป็นพารามิเตอร์ แล้วฟังก์ชันมวลความน่าจะเป็นของ X แสดงดังสมการที่ (11)

$$f(x) = \frac{\theta^2}{(\theta+1)^{x+2}} \left(1 + \frac{\alpha+\beta x}{\theta\alpha+\beta}\right), x = 0, 1, 2, \dots \quad (11)$$

โดยที่ $\theta > 0, \beta > 0$ และ $\theta\alpha + \beta > 0$

ค่าเฉลี่ยและความแปรปรวนสำหรับตัวแปรสุ่ม แสดงดังสมการที่ (12) และ (13) ตามลำดับ

$$E(X) = \frac{\theta\alpha+2}{\theta(\theta\alpha+\beta)} \quad (12)$$

$$Var(X) = \frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)} + \frac{2(\theta\alpha+3\beta)}{\theta^2(\theta\alpha+\beta)} - \left\{ \frac{\theta\alpha+2\beta}{\theta(\theta\alpha+\beta)} \right\}^2 \quad (13)$$

ดัชนีการกระจายตัว แสดงดังสมการที่ (14)

$$ID = 1 + \frac{\theta^2+4\theta\alpha\beta+2\beta^2}{\theta(\theta\alpha+\beta)(\theta\alpha+2\beta)} \quad (14)$$

การประมาณค่าพารามิเตอร์ของตัวอย่างสุ่ม $X_1, X_2, \dots, X_n$ ที่ได้จากการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ ดำเนินการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด ซึ่งมีฟังก์ชันล๊อคภาวะน่าจะเป็นดังสมการที่ (15)

$$\ln L = n \ln \left(\frac{\theta^2}{\theta\alpha + \beta} \right) - \sum_{i=1}^n (x_i + 2) \log(\theta + 1) + \sum_{i=1}^n \log(\theta\alpha + \beta + \alpha + \beta x_i) \quad (15)$$

2.4 เกณฑ์สำหรับการเปรียบเทียบ

ในการวิจัยนี้ได้ศึกษาการเปรียบเทียบการแจกแจงแบบปัวซอง-ลินด์เลย์แบบทั่วไป การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ โดยใช้เกณฑ์ดังนี้

1) ค่าล๊อคภาวะน่าจะเป็น คือ ลอการิทึมของฟังก์ชันภาวะน่าจะเป็นซึ่งใช้ในการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด กำหนดให้ $L(\theta)$ แทนฟังก์ชันภาวะน่าจะเป็นของการแจกแจง

และ $\ell(\theta)$ แทนลอการิทึมของฟังก์ชันความน่าจะเป็น ซึ่งสามารถเขียนได้แสดงดังสมการที่ (16) (Akaike, 1974)

$$\ell(\theta) = \sum_{i=1}^n \ln f(x_i|\theta) \quad (16)$$

เมื่อ $f(x_i|\theta)$ แทนฟังก์ชันความหนาแน่นหรือฟังก์ชันมวลความน่าจะเป็นของข้อมูลตัวอย่าง x_i ภายใต้พารามิเตอร์ θ และ n คือขนาดตัวอย่าง

2) เกณฑ์สารสนเทศของอะกะอิเกะ เป็นหนึ่งในเครื่องมือสำคัญสำหรับการประเมินและเลือกแบบจำลองทางสถิติ โดยคำนึงถึงทั้งความพอดี (Goodness of fit) และความซับซ้อนของการแจกแจง เพื่อหลีกเลี่ยงปัญหาการประมาณค่าที่เกินจริง สมมติให้ข้อมูลชุดหนึ่งมีการแจกแจงทางสถิติหลายแบบ ซึ่งแต่ละการแจกแจงมีจำนวนพารามิเตอร์ที่แตกต่างกันไป ค่า AIC สามารถคำนวณได้โดยใช้สูตรแสดงดังสมการที่ (17) (Akaike, 1974; Burnham & Anderson, 2002)

$$AIC = -2 \ln(L) + 2k \quad (17)$$

เมื่อ k คือ จำนวนพารามิเตอร์ในตัวแบบ และ $\ln(L)$ คือค่าลอการิทึมความน่าจะเป็นที่ประเมินจากค่าประมาณพารามิเตอร์

3) เกณฑ์สารสนเทศของเบย์เป็นเกณฑ์สำหรับการเลือกแบบจำลองทางสถิติที่เสนอโดย Schwarz (1978) ซึ่งมีพื้นฐานจากกรอบคิดแบบเบย์ โดยประเมินความพอดีของการแจกแจงขึ้นอยู่กับจำนวนพารามิเตอร์และขนาดตัวอย่างสามารถคำนวณได้โดยใช้สูตรแสดงดังสมการที่ (18) (Burnham & Anderson, 2002)

$$BIC = -2 \ln(L) + k \ln(n) \quad (18)$$

เมื่อ n คือ ขนาดของตัวอย่าง k คือ จำนวนพารามิเตอร์ในตัวแบบ และ $\ln(L)$ คือ ค่าลอการิทึมความน่าจะเป็นที่ประเมินจากค่าประมาณพารามิเตอร์

4) การทดสอบคอลโมโกรอฟ-สมิรโนฟ ในการตรวจสอบความเหมาะสมของแบบจำลองการแจกแจงได้ใช้การทดสอบคอลโมโกรอฟ-สมิรโนฟสำหรับข้อมูลไม่ต่อเนื่อง โดยดำเนินการผ่านฟังก์ชัน $ks.test()$ จากแพ็คเกจ $dgof$ ในโปรแกรม R ซึ่งเป็นการปรับปรุงจากการทดสอบ KS แบบมาตรฐานที่ใช้กับข้อมูลต่อเนื่องเพื่อให้เหมาะสมกับข้อมูลเชิงนับ การทดสอบนี้พิจารณาความแตกต่างสูงสุดระหว่างฟังก์ชันการแจกแจงเชิงประจักษ์ (Empirical distribution function) ของข้อมูลกับฟังก์ชันความหนาแน่นสะสม (Cumulative distribution function) ของการแจกแจงโดยสถิติทดสอบสามารถเขียนได้แสดงดังสมการที่ (19)

$$D = \sup_x |F_0(x) - F_{data}(x)| \quad (19)$$

โดย $F_{data}(x)$ คือ ฟังก์ชันการแจกแจงเชิงประจักษ์ของข้อมูลตัวอย่าง และ $F_0(x)$ คือฟังก์ชันการแจกแจงสะสมของแบบจำลองที่พิจารณา ค่าที่ของสถิติทดสอบคำนวณด้วยวิธีการจำลองมอนติคาร์โล (Monte carlo simulation) ตามขั้นตอนของแฟกเกดดังกล่าว การทดสอบนี้สำหรับข้อมูลไม่ต่อเนื่องสามารถใช้กับข้อมูลเชิงนับได้โดยไม่ต้องจัดกลุ่มข้อมูลต่างจากการทดสอบไคกำลังสองของเพียร์-สันซึ่งต้องมีการจัดกลุ่มข้อมูลและมีข้อกำหนดเกี่ยวกับค่าคาดหวังในแต่ละกลุ่ม โดยมีสมมติฐานการทดสอบดังนี้

$$\begin{aligned}H_0: F_{data}(x) &= F_0(x) \\H_1: F_{data}(x) &\neq F_0(x)\end{aligned}$$

หากปฏิเสธสมมติฐาน H_0 สรุปได้ว่า แบบจำลองดังกล่าวไม่เหมาะสมกับข้อมูล (Arnold & Emerson, 2011)

2.5 ความถี่คาดหวัง

ในการประเมินความเหมาะสมของการแจกแจงข้อมูลเชิงนับสามารถพิจารณาความถี่คาดหวัง (Expected frequencies) ที่ได้จากฟังก์ชันมวลความน่าจะเป็นของการแจกแจงที่กำหนด โดยกำหนดให้ $p(x; \theta)$ แทนค่าความน่าจะเป็นของตัวแปรสุ่มที่มีค่าเท่ากับ x ภายใต้พารามิเตอร์ θ และ n แทนขนาดตัวอย่าง ความถี่คาดหวังสำหรับค่าข้อมูล x สามารถคำนวณได้จากสมการที่ (20)

$$E_x = n \cdot p(x; \theta) \quad (20)$$

โดยที่ E_x คือ ความถี่คาดหวังของค่าข้อมูล x จากแบบจำลอง ซึ่งการเปรียบเทียบความถี่คาดหวังและความถี่ที่สังเกตได้ช่วยให้สามารถประเมินความสอดคล้องระหว่างการแจกแจงกับข้อมูลจริงได้ หากความถี่คาดหวังมีค่าใกล้เคียงกับความถี่ที่สังเกตได้ในแต่ละค่าของข้อมูล หมายความว่าแบบจำลองนั้นน่าจะเหมาะสมที่จะอธิบายข้อมูลได้อย่างเหมาะสม แนวคิดดังกล่าวนี้นักใช้ประกอบการทดสอบความเหมาะสมของการแจกแจงในการวิเคราะห์ข้อมูลเชิงนับ ในงานวิจัยนี้ความถี่คาดหวังคำนวณจากฟังก์ชันมวลความน่าจะเป็นของการแจกแจงที่ใช้เปรียบเทียบโดยค่าพารามิเตอร์ของแต่ละการแจกแจงประมาณด้วยวิธี MLE จากนั้นนำค่าพารามิเตอร์ที่ได้ไปคำนวณความถี่คาดหวังสำหรับแต่ละค่าของข้อมูล และเปรียบเทียบกับความถี่ที่สังเกตได้ เพื่อใช้ประกอบการประเมินความเหมาะสมของแบบจำลองกับข้อมูลเชิงประจักษ์ (Casella & Berger, 2002)

2.6 การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ของการวิจัยนี้ดำเนินการโดยใช้วิธีภาวะน่าจะเป็นสูงสุด ซึ่งเป็นวิธีที่นิยมใช้กันอย่างแพร่หลาย โดยมีหลักการ คือ กำหนดให้ $Y_1, Y_2, \dots, Y_n$ เป็นตัวอย่างสุ่มที่มีความเป็นอิสระต่อกัน และมีฟังก์ชันความน่าจะเป็น $f(y_i|\theta)$ ฟังก์ชันภาวะน่าจะเป็นสามารถเขียนได้เป็น $L(\theta) = \prod_{i=1}^n f(y_i|\theta)$ และสามารถพิจารณาในรูปของฟังก์ชันลอการิทึมภาวะน่าจะเป็น

$\ell(\theta) = \sum_{i=1}^n \log(f(y_i|\theta))$ เพื่อความสะดวกในการคำนวณ โดยค่าประมาณของพารามิเตอร์จะได้จากค่าสูงสุดของฟังก์ชันล๊อกลักษณะน่าจะเป็น (Casella & Berger, 2002) ในกรณีที่ตัวประมาณไม่ได้อยู่ในรูปแบบชัดแจ้ง (Closed form) สามารถใช้วิธีการหาค่าสูงสุดโดยการหาค่าประมาณเชิงตัวเลข (Numerical optimization) ซึ่งในการวิจัยนี้การคำนวณค่าประมาณพารามิเตอร์ดำเนินการโดยใช้ฟังก์ชัน optim() ในโปรแกรม R (R Core Team, 2024) สำหรับการหาค่าประมาณพารามิเตอร์ที่ไม่ทราบค่า และการหาค่าสูงสุดของฟังก์ชันลักษณะน่าจะเป็นใช้วิธี Broyden Fletcher Goldfarb Shanno algorithm (BFGS)

ในการวิจัยนี้ค่าเริ่มต้นของพารามิเตอร์ถูกกำหนดโดยอาศัยข้อมูลจริงของชุดข้อมูล เช่น ค่าเฉลี่ยตัวอย่าง ซึ่งค่าพารามิเตอร์ของการแจกแจงสำหรับการวิจัยนี้ต้องมีค่าเป็นจำนวนจริงบวก เพื่อให้ฟังก์ชันความน่าจะเป็นมีความถูกต้องตามทฤษฎีบท ดังนั้น ในการประมาณค่าพารามิเตอร์จึงกำหนดเงื่อนไขว่า $\theta > 0$ เมื่อ θ คือ พารามิเตอร์ที่ไม่ทราบค่า โดยใช้วิธีการกำหนดขอบเขตล่างของพารามิเตอร์หาค่าประมาณเชิงตัวเลข แต่ในบางกรณีค่าประมาณของพารามิเตอร์อาจมีค่าที่ใกล้ศูนย์มาก ๆ ซึ่งอาจก่อให้เกิดปัญหาความไม่เสถียรของการคำนวณเชิงตัวเลข ดังนั้นในการคำนวณจึงกำหนดค่าเกณฑ์ขั้นต่ำ โดยค่าที่ต่ำกว่าค่านี้จะถูกพิจารณาว่าเข้าใกล้ศูนย์แต่ยังคงอยู่ในขอบเขตของพารามิเตอร์ที่ยอมรับได้ และในการประมาณค่าพารามิเตอร์จะทำการตรวจสอบการลู่เข้าของกระบวนการหาค่าสูงสุดจากสถานะการลู่เข้าที่รายงานโดยฟังก์ชัน optim() และตรวจสอบว่าค่าของฟังก์ชันความน่าจะเป็นล๊อกลักษณะน่าจะเป็นก่อนนำผลลัพธ์ไปใช้ในการเปรียบเทียบแบบจำลอง

2.7 ชุดข้อมูล

ชุดข้อมูลที่นำมาใช้ในการประยุกต์และเปรียบเทียบในการวิจัยนี้ประกอบด้วย ข้อมูลจำนวน 3 ชุด ได้แก่ ข้อมูลชุดที่ 1 คือ จำนวนผู้เสียชีวิตจำแนกตามสาเหตุอุบัติเหตุทางถนนของจังหวัดระยอง ระหว่างปี พ.ศ. 2560-2567 ซึ่งได้มาจากเว็บไซต์ศูนย์กลางข้อมูลเปิดภาครัฐ (<https://gdcatalog.go.th/dataset/>) โดยมีค่าเฉลี่ยเท่ากับ 1.0417 ค่าความแปรปรวนเท่ากับ 1.8310 และค่าดัชนีการกระจายเท่ากับ 1.7577 ข้อมูลชุดที่ 2 เป็นจำนวนผู้เสียชีวิตจากโรคติดเชื้อ COVID-19 ของจังหวัดศรีสะเกษ ระหว่างปี พ.ศ. 2563-2567 ซึ่งได้มาจากแหล่งข้อมูลเดียวกัน โดยมีค่าเฉลี่ยเท่ากับ 4.3977 ค่าความแปรปรวนเท่ากับ 48.8170 และค่าดัชนีการกระจายเท่ากับ 11.1005 และข้อมูลชุดที่ 3 เป็นข้อมูลทุติยภูมิจากแพ็กเกจ AER ในโปรแกรม R (Kleiber & Zeileis, 2025) ซึ่งเป็นข้อมูลจำนวนครั้งในการพบแพทย์ภายในช่วงเวลา 2 สัปดาห์ จากการสำรวจสุขภาพประชากรประเทศออสเตรเลีย ระหว่างปี ค.ศ. 1977-1978 โดยมีค่าเฉลี่ยเท่ากับ 0.3017 ค่าความแปรปรวนเท่ากับ 0.6370 และค่าดัชนีการกระจายเท่ากับ 2.1114 ทั้งนี้ ข้อมูลทั้งสามชุดมีค่าดัชนีการกระจายมากกว่า 1 ซึ่งบ่งชี้ว่า ข้อมูลมีลักษณะการกระจายเกินเกณฑ์และเหมาะสมสำหรับการประเมินประสิทธิภาพของการแจกแจงปัวซองผสมที่พิจารณาในการศึกษานี้

3. ผลการวิจัยและอภิปรายผล

3.1 ผลการวิจัย

ในส่วนนี้นำเสนอผลการเปรียบเทียบประสิทธิภาพของการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ และยังได้ทำการ

เปรียบเทียบผลลัพธ์กับการแจกแจงปัวซอง และการแจกแจงทวินามลบ ผ่านการประยุกต์ใช้กับชุดข้อมูลจริงซึ่งมีลักษณะการกระจายเกินเกณฑ์ เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของการแจกแจงทั้ง 5 การแจกแจง ได้แก่ ค่า -LL AIC และ BIC นอกจากนี้ ยังได้ทำการทดสอบความเหมาะสมของการแจกแจงโดยใช้สถิติทดสอบ KS ที่ระดับนัยสำคัญ 0.05 ควบคู่กับการคำนวณความถี่ค่าความหมายโดยสรุปผลการวิเคราะห์ไว้ในตารางที่ 1 ถึงตารางที่ 6

ตารางที่ 1 แสดงความถี่ค่าความหมายและค่าประมาณพารามิเตอร์สำหรับข้อมูลชุดที่ 1 ในขณะที่ตารางที่ 2 แสดงผลการเปรียบเทียบความเหมาะสมของการแจกแจง โดยพิจารณาจากสถิติการทดสอบ KS รวมทั้งค่า -LL AIC และ BIC ผลการทดสอบความเหมาะสมด้วยการทดสอบ KS แสดงให้เห็นว่า การแจกแจงทั้ง 5 แบบ ให้ความสอดคล้องกับข้อมูลที่ระดับนัยสำคัญ 0.05 ซึ่งสอดคล้องกับลักษณะของข้อมูลที่มีค่าดัชนีการกระจายมากกว่า 1 อันบ่งชี้ถึงการกระจายเกินเกณฑ์ เมื่อพิจารณาเกณฑ์การเลือกแบบจำลองพบว่า การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ค่า -LL AIC และ BIC ต่ำที่สุด ขณะที่ค่าดังกล่าวมีความใกล้เคียงกับแบบจำลองทวินามลบ ซึ่งเป็นการแจกแจงมาตรฐานสำหรับข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์ แสดงให้เห็นว่า การแจกแจงทั้งสองสามารถอธิบายข้อมูลได้ในระดับที่ใกล้เคียงกัน นอกจากนี้ ความถี่ค่าความหมายที่ได้จากการแจกแจงดังกล่าวมีค่าใกล้เคียงกับความถี่ที่สังเกตได้ในเกือบทุกค่าของข้อมูล ดังแสดงในรูปที่ 1

ตารางที่ 3 แสดงความถี่ค่าความหมายและค่าประมาณพารามิเตอร์สำหรับการแจกแจงทั้ง 5 การแจกแจงสำหรับข้อมูลชุดที่ 2 พบว่า การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดค่าประมาณบางค่าอาจมีค่าใกล้ศูนย์มาก เช่น 0.000001 อาจเกิดจากข้อจำกัดของการคำนวณเชิงตัวเลข ค่าที่มีขนาดเล็กมากดังกล่าวจึงสามารถตีความได้ว่า มีค่าใกล้ศูนย์ในเชิงปฏิบัติ แต่ไม่ได้ส่งผลกระทบต่อคุณสมบัติหลักของการแจกแจง และตารางที่ 4 แสดงผลการวิเคราะห์ข้อมูลชุดที่ 2 พบว่า ผลการทดสอบความเหมาะสมด้วยการทดสอบ KS การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก การแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ และการแจกแจงทวินามลบ ให้ความสอดคล้องกับข้อมูลที่ระดับนัยสำคัญ 0.05 และเมื่อพิจารณาเกณฑ์การเปรียบเทียบการแจกแจง พบว่า การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป ให้ค่า -LL, AIC และ BIC ใกล้เคียงกับการแจกแจงทวินามลบ ซึ่งแสดงให้เห็นว่าการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ผลดีกว่าการแจกแจงที่เปรียบเทียบสำหรับข้อมูลชุดที่ 2 ทั้งนี้ผลการประมาณความถี่ค่าความหมายจากการแจกแจงสอดคล้องกับข้อมูลที่สังเกตได้ ดังแสดงในรูปที่ 2

ตารางที่ 1. ความถี่ค่าสังเกต ความถี่ค่าความหมาย และค่าประมาณพารามิเตอร์ของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 1

จำนวนผู้เสียชีวิต จำแนกตามสาเหตุอุบัติเหตุ	ความถี่ค่าสังเกต	ความถี่ค่าความหมายของการแจกแจง				
		Poisson	GPL	PWL	TPPLD	NB
0	111	84.69	111.13	110.94	105.49	110.82
1	65	88.22	64.46	65.18	64.52	65.27

ตารางที่ 1. (ต่อ) ความถี่ค่าสังเกต ความถี่ค่าคาดหวัง และค่าประมาณพารามิเตอร์ของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 1

จำนวน ผู้เสียชีวิต จำแนกตาม สาเหตุ อุบัติเหตุ	ความถี่ ค่าสังเกต	ความถี่ค่าคาดหวังของการแจกแจง				
		Poisson	GPL	PWL	TPPLD	NB
2	33	45.95	33.61	33.25	35.19	33.40
3	17	15.95	16.51	16.23	18.02	16.24
4	7	4.16	7.80	7.72	8.87	7.68
5	4	0.87	3.59	3.61	4.24	3.57
6	2	0.15	1.62	1.66	1.99	1.65
7	0	0.02	0.72	0.76	0.92	1.03
8	1	0.00	0.32	0.34	0.42	0.34
รวม	240	240	240	240	240	240
ค่าประมาณพารามิเตอร์		$\hat{\lambda} = 1.0417$	$\hat{\alpha} = 2.1999$	$\hat{\theta} = 10.3208$	$\hat{\alpha} = 0.6528$	$\hat{r} = 1.3552$
		-	$\hat{\theta} = 1.5267$	$\hat{\alpha} = 1.4290$	$\hat{\theta} = 1.4359$	$\hat{p} = 0.5654$
		-	-	-	$\hat{\beta} = 1.5267$	-

ตารางที่ 2. เกณฑ์การเปรียบเทียบของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 1

เกณฑ์การเปรียบเทียบ	การแจกแจง				
	Poisson	GPL	PWL	TPPLD	NB
จำนวนพารามิเตอร์	1	2	2	3	2
- LL	358.2877	338.7086	338.7123	338.7086	338.7179
AIC	718.5754	681.4172	681.4246	683.4172	681.4358
BIC	722.0561	688.3785	688.3859	693.8592	688.3971
KS-test	0.4092	0.3360	0.3791	0.3595	0.3599
P-value	0.0621	0.1877	0.1034	0.1385	0.1390

ตารางที่ 5 แสดงแสดงความถี่ค่าคาดหวังและค่าประมาณพารามิเตอร์ของการแจกแจงทั้ง 5 แบบ สำหรับข้อมูลชุดที่ 3 และตารางที่ 6 แสดงผลการวิเคราะห์ข้อมูลชุดที่ 3 ซึ่งพบว่า จากผลการทดสอบ KS การแจกแจงทวินามลบให้ค่าพีมากกว่าระดับนัยสำคัญ 0.05 และการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ค่าพี 0.0501 จึงยังไม่พบหลักฐานเพียงพอที่จะปฏิเสธความเหมาะสมของแบบจำลองที่ระดับนัยสำคัญ 0.05 นอกจากนี้ เมื่อพิจารณาเกณฑ์การเปรียบเทียบการแจกแจง พบว่า การแจกแจงทวินามลบให้ค่า -LL, AIC และ BIC ต่ำที่สุด รองลงมาคือการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป แสดงให้เห็นว่าการแจกแจง

ดังกล่าวสามารถแสดงศักยภาพในการอธิบายข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ เมื่อเปรียบเทียบกับแบบจำลองการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนักและการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ ดังแสดงในรูปที่ 3

ตารางที่ 3. ความถี่ค่าสังเกต ความถี่คาดหวัง และค่าประมาณพารามิเตอร์ของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 2

จำนวน ผู้เสียชีวิต COVID-19	ความถี่ ค่าสังเกต	ความถี่คาดหวังของการแจกแจง				
		Poisson	GPL	PWL	TPPLD	NB
0	54	3.26	28.52	7.76	24.34	43.39
1	13	11.48	21.88	12.46	18.96	13.09
2	7	20.19	16.44	14.30	14.76	8.01
3	4	23.67	12.17	14.13	11.49	5.72
4	3	20.82	8.89	12.80	8.95	4.39
5	4	14.65	6.43	10.95	6.97	3.51
6	4	8.59	4.62	8.99	5.43	2.87
7	3	4.32	3.29	7.16	4.23	2.39
8	4	1.90	2.33	5.57	3.29	2.02
9	1	0.74	1.64	4.24	2.56	1.72
10	0	0.26	1.15	3.19	2.00	1.56
11	2	0.08	0.81	2.36	1.55	1.28
12	3	0.02	0.56	1.73	1.21	1.11
13	1	0.01	0.39	1.26	0.94	0.98
14	1	0.00	0.27	0.91	0.73	0.86
15	2	0.00	0.19	0.65	0.57	0.75
21	2	0.00	0.02	0.08	0.13	0.37
31	1	0.00	0.00	0.00	0.01	0.13
⋮	⋮	⋮	⋮	⋮	⋮	⋮
42	1	0.00	0.00	0.00	0.00	0.94
รวม	110	110	110	110	110	110
ค่าประมาณพารามิเตอร์		$\hat{\lambda} = 3.5182$	$\hat{\alpha} = 0.000001$	$\hat{\alpha} = 0.3009$	$\hat{\alpha} = 3.8292$	$\hat{r} = 0.3272$
		-	$\hat{\theta} = 0.2273$	$\hat{\theta} = 0.5194$	$\hat{\theta} = 0.2843$	$\hat{p} = 0.0779$
		-	-	-	$\hat{\beta} = 0.000001$	-

ตารางที่ 4. เกณฑ์การเปรียบเทียบของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 2

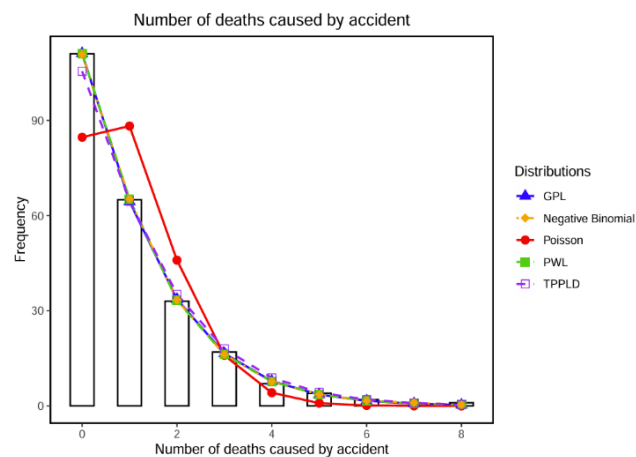
เกณฑ์การเปรียบเทียบ	การแจกแจง				
	Poisson	GPL	PWL	TPPLD	NB
จำนวนพารามิเตอร์	1	2	2	3	2
-LL	533.1240	227.6585	265.2079	262.706	227.1118
AIC	1068.2480	459.3170	534.4158	531.4121	458.2236
BIC	1070.9490	464.7180	539.8168	539.5134	463.6246
KS-test	0.3181	0.2273	0.2778	0.2222	0.2222
P-value	0.0206	0.1652	0.0522	0.1652	0.1975

ตารางที่ 5. ความถี่ค่าสังเกต ความถี่ค่าคาดหวัง และค่าประมาณพารามิเตอร์ของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 3

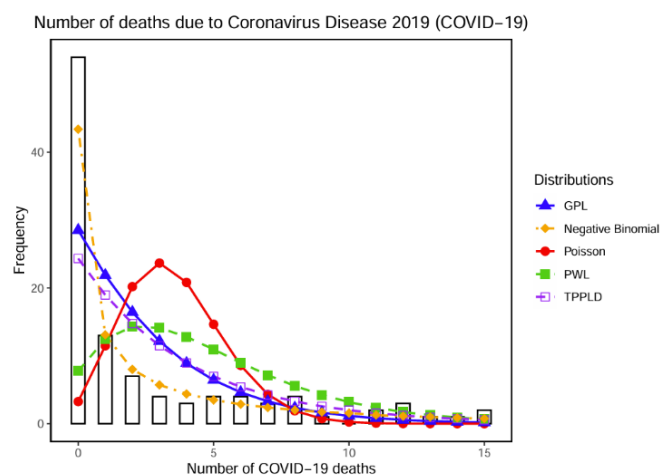
จำนวนครั้ง พบแพทย์	ความถี่ ค่าสังเกต	ความถี่ค่าคาดหวังของการแจกแจง				
		Poisson	GPL	PWL	TPPLD	NB
0	4141	3838.19	3986.99	3931.89	4509.64	4157.80
1	782	1158.11	924.16	1012.93	591.17	697.28
2	174	174.72	214.21	201.20	77.50	213.33
3	30	17.57	49.65	36.40	10.16	75.10
4	24	1.33	11.51	6.30	1.33	28.17
5	9	0.08	2.67	1.07	0.17	10.95
6	12	0.00	0.62	0.18	0.02	4.36
7	12	0.00	0.14	0.03	0.00	1.76
8	5	0.00	0.03	0.01	0.00	0.72
9	1	0.00	0.01	0.01	0.00	0.30
รวม	5190	5190	5190	5190	5190	5190
ค่าประมาณพารามิเตอร์		$\hat{\lambda} = 0.3017$	$\hat{\alpha} = -0.000001$	$\hat{\alpha} = 1.0148$	$\hat{\alpha} = 1.9923$	$\hat{r} = 0.3776$
		-	$\hat{\theta} = 3.3141$	$\hat{\theta} = 5.5281$	$\hat{\theta} = 0.000001$	$\hat{p} = 0.5558$
		-	-	-	$\hat{\beta} = 6.6284$	-

ตารางที่ 6. เกณฑ์การเปรียบเทียบของแต่ละการแจกแจงสำหรับข้อมูลชุดที่ 3

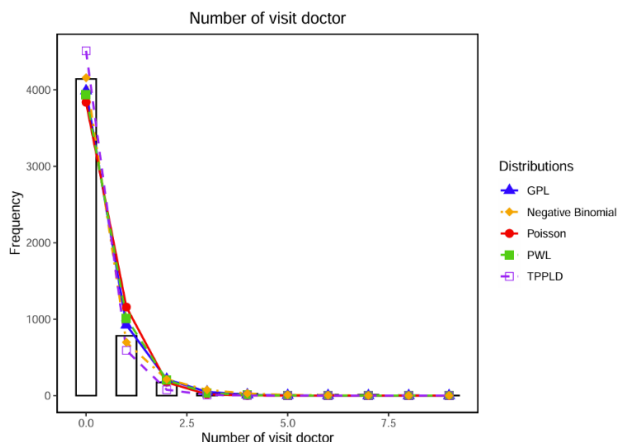
เกณฑ์การเปรียบเทียบ	การแจกแจง				
	Poisson	GPL	PWL	TPPLD	NB
จำนวนพารามิเตอร์	1	2	2	3	2
-LL	3983.1940	3657.9340	3744.2600	3657.9321	3585.9920
AIC	7968.3890	7319.8680	7492.5200	7321.8642	7175.9840
BIC	7974.9425	7332.9770	7505.6290	7341.5277	7189.0930
KS-test	0.5268	0.4024	0.4000	0.4571	0.400
P-value	0.0035	0.0501	0.0427	0.0161	0.0517



รูปที่ 1. กราฟความถี่ค่าสังเกตและความถี่คาดหวังสำหรับการแจกแจงต่าง ๆ สำหรับข้อมูลชุดที่ 1



รูปที่ 2. กราฟความถี่ค่าสังเกตและความถี่คาดหวังสำหรับการแจกแจงต่าง ๆ สำหรับข้อมูลชุดที่ 2



รูปที่ 3. กราฟความถี่ค่าสังเกตและความถี่คาดหวังสำหรับการแจกแจงต่าง ๆ สำหรับข้อมูลชุดที่ 3

3.2 อภิปรายผล

จากผลการวิเคราะห์ข้อมูลจริงทั้ง 3 ชุด พบว่า การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป แสดงให้เห็นถึงประสิทธิภาพในการอธิบายข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ได้ดีกว่าการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ โดยให้ค่า -LL AIC และ BIC ต่ำกว่าทั้งสองการแจกแจง อย่างไรก็ตามความแตกต่างของค่า AIC และ BIC ระหว่างการแจกแจงดังกล่าวกับการแจกแจงทวินามลบมีค่าไม่มาก ซึ่งบ่งชี้ว่าการแจกแจงทั้งสองมีระดับความเหมาะสมกับข้อมูลใกล้เคียงกัน นอกจากนี้ผลการทดสอบความเหมาะสมด้วยสถิติ KS สำหรับทั้ง 3 ชุดข้อมูล การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ค่าที่มากกว่าระดับนัยสำคัญ 0.05 สะท้อนให้เห็นว่าการแจกแจงนี้สามารถอธิบายโครงสร้างของข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ได้ดีอีกสองการแจกแจง แม้ว่าการเพิ่มพารามิเตอร์ในบางการแจกแจงจะช่วยเพิ่มความยืดหยุ่นเชิงทฤษฎี แต่ผลการศึกษาเชิงประยุกต์ในงานวิจัยนี้ชี้ให้เห็นว่าการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป ยังคงเป็นทางเลือกที่มีประสิทธิภาพและเหมาะสมสำหรับการวิเคราะห์ข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์

ผลการวิเคราะห์ในงานวิจัยนี้สอดคล้องกับงานวิจัยก่อนหน้านี้ที่ศึกษาแบบจำลองในตระกูลการแจกแจงปัวซอง-ลินด์เลย์ซึ่งได้รายงานว่า การแจกแจงที่พัฒนาจากการแจกแจงพื้นฐานดังกล่าวสามารถเพิ่มความยืดหยุ่นในการอธิบายข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ได้ดีขึ้น เช่น งานของ Bhati et al. (2015) และ Atikankul et al. (2020) ที่เสนอการปรับปรุงรูปแบบของการแจกแจงปัวซอง-ลินด์เลย์เพื่อเพิ่มความสามารถในการรองรับความแปรปรวนของข้อมูล ในทำนองเดียวกัน Das et al. (2018) ได้เสนอรูปแบบการแจกแจงที่มีพารามิเตอร์เพิ่มเติมเพื่อเพิ่มความยืดหยุ่นของแบบจำลองสำหรับข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์ จากผลการศึกษาพบว่า การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ค่าของ -LL AIC และ BIC ต่ำกว่าการแจกแจงอื่นสำหรับข้อมูลที่ศึกษา ซึ่งแสดงให้เห็นว่า การแจกแจงดังกล่าวสามารถอธิบายลักษณะการกระจายของข้อมูลได้ดีกว่า การที่การแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ผลการประมาณที่ดีกว่าอาจเนื่องมาจากโครงสร้างของพารามิเตอร์ที่ช่วยเพิ่มความยืดหยุ่นในการควบคุมทั้ง

ค่าเฉลี่ยและความแปรปรวนของข้อมูล ทำให้สามารถรองรับข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์ได้อย่างเหมาะสม เมื่อเปรียบเทียบกับแบบจำลองการแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนักและการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ ซึ่งแม้จะมีพารามิเตอร์เพิ่มเติมเช่นกัน แต่ยังไม่สามารถสะท้อนรูปแบบการกระจายของข้อมูลจริงในชุดข้อมูลที่ศึกษานี้ได้ดีเท่ากับการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป

อย่างไรก็ตาม ผลการศึกษายังแสดงให้เห็นว่าการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป และการแจกแจงทวินามลบซึ่งเป็นแบบจำลองมาตรฐานสำหรับข้อมูลเชิงนับที่มีการกระจายเกินเกณฑ์ ให้ค่าของ AIC และ BIC ที่ใกล้เคียงกัน ซึ่งบ่งชี้ว่าแบบจำลองทั้งสองมีระดับความเหมาะสมกับข้อมูลใกล้เคียงกันในเชิงปฏิบัติ นอกจากนี้ งานวิจัยนี้ยังมีข้อจำกัดบางประการ กล่าวคือ การวิเคราะห์ที่ใช้ข้อมูลจริงเพียง 3 ชุดข้อมูล ซึ่งอาจยังไม่เพียงพอสำหรับการสรุปเชิงทั่วไปเกี่ยวกับประสิทธิภาพของแบบจำลองในทุกสถานการณ์

4. สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไป การแจกแจงปัวซอง-ลินด์เลย์แบบถ่วงน้ำหนัก และการแจกแจงปัวซอง-ลินด์เลย์สามพารามิเตอร์ สำหรับการวิเคราะห์ข้อมูลเชิงนับที่มีลักษณะการกระจายเกินเกณฑ์ โดยใช้ข้อมูลจริงจำนวน 3 ชุด ได้แก่ ข้อมูลจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในจังหวัดระยอง ข้อมูลจำนวนผู้เสียชีวิตจากโรคติดเชื้อ COVID-19 ในจังหวัดศรีสะเกษ และข้อมูลจำนวนครั้งในการพบแพทย์ภายในระยะเวลา 2 สัปดาห์ ผลการวิเคราะห์โดยใช้ค่า -LL AIC BIC และการทดสอบ KS สำหรับข้อมูลไม่ต่อเนื่อง พบว่า การแจกแจงในตระกูลปัวซอง-ลินด์เลย์ และการแจกแจงทวินามลบสามารถอธิบายข้อมูลที่มีลักษณะการกระจายเกินเกณฑ์ได้ดีกว่าการแจกแจงปัวซอง และการแจกแจงปัวซอง-ลินด์เลย์แบบทั่วไปให้ค่าของเกณฑ์การเลือกแบบจำลองที่ดีกว่าในหลายกรณีซึ่งบ่งชี้ว่าการแจกแจงดังกล่าวมีความเหมาะสมกับข้อมูลในเชิงประจักษ์ อย่างไรก็ตาม ความแตกต่างของค่า AIC และ BIC ระหว่างการแจกแจงบางแบบมีค่าไม่มาก จึงอาจกล่าวได้ว่าการแจกแจงหลายแบบยังคงมีระดับความเหมาะสมกับข้อมูลใกล้เคียงกัน อย่างไรก็ตามการศึกษานี้ยังมีข้อจำกัดบางประการ กล่าวคือการวิเคราะห์ที่ใช้ข้อมูลจริงเพียง 3 ชุดข้อมูล และการเปรียบเทียบยังจำกัดอยู่ในกลุ่มการแจกแจงปัวซอง-ลินด์เลย์ และการแจกแจงทวินามลบ ดังนั้น งานวิจัยในอนาคตอาจขยายการศึกษาไปยังชุดข้อมูลที่มีลักษณะหลากหลายมากขึ้น รวมถึงการเปรียบเทียบกับแบบจำลองอื่นสำหรับข้อมูลเชิงนับ นอกจากนี้ แนวทางการศึกษาต่ออาจพัฒนาแบบจำลองถดถอยเพื่อใช้ในการวิเคราะห์ข้อมูลที่มีตัวแปรอธิบาย หรือพิจารณาการขยายแบบจำลองไปสู่ตัวแบบค่าศูนย์เพื่อ (Zero-inflated model) ซึ่งเหมาะสมสำหรับข้อมูลเชิงนับที่มีจำนวนศูนย์มากกว่าที่แบบจำลองมาตรฐานสามารถอธิบายได้

เอกสารอ้างอิง (References)

- Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M., & Eliwa, M. S. (2022). Poisson XLindley distribution for count data: Statistical and reliability properties with estimation techniques and inference. *Computational Intelligence and Neuroscience*, 2022(1), Article 6503670. <https://doi.org/10.1155/2022/6503670>

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alomair, G., Tajuddin, R. R. M., Bakouch, H. S., & Almohisen, A. (2024). A statistical model for count data analysis and population size estimation: Introducing a mixed Poisson-Lindley distribution and its zero truncation. *Axioms*, 13(2), Article 125. <https://doi.org/10.3390/axioms13020125>
- Arnold, T. B., & Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3(2), 34-39. <https://doi.org/10.32614/RJ-2011-016>
- Atikankul, Y. (2023). A new poisson generalized Lindley regression model. *Austrian Journal of Statistics*, 52(1), 39-50. <https://doi.org/10.17713/ajs.v52i1.1344>
- Atikankul, Y., Thongteeraparp, A., & Bodhisuwan, W. (2020). The new Poisson mixed weighted Lindley distribution with applications to insurance claims data. *Songklanakarin Journal of Science and Technology*, 42(1), 152-162. <https://doi.org/10.14456/sjst-psu.2020.21>
- Bhati, D., Sastry, D. V. S., & Qadri, P. Z. M. (2015). A new generalized Poisson-Lindley distribution: Applications and properties. *Austrian Journal of Statistics*, 44(4), 35-51. <https://doi.org/10.17713/ajs.v44i4.54>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer New York. <https://doi.org/10.1007/b97636>
- Cameron, A. C., & Trivedi, P. K. (2014). *Regression analysis of count data* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013567>
- Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury Press.
- Das, K. K., Ahmed, I., & Bhattacharjee, S. (2018). A new three-parameter Poisson-Lindley distribution for modelling over-dispersed count data. *International Journal of Applied Engineering Research*, 13(23), 16468-16477. <https://doi.org/10.37622/IJAER/13.23.2024.16468-16477>
- Kleiber, C., & Zeileis, A. (2025). *AER: Applied econometrics with R (R package version 1.2-15)*. Comprehensive R Archive Network (CRAN). <https://CRAN.R-project.org/package=AER>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464. <https://doi.org/10.1214/aos/1176344136>

- Tharshan, R., & Wijekoon, P. (2022a). Zero-modified Poisson-modification of Quasi Lindley distribution and its application. *Statistics in Transition New Series*, 23(4), 113-128.
<https://doi.org/10.2478/stattrans-2022-0045>
- Tharshan, R., & Wijekoon, P. (2022b). A new mixed Poisson distribution for over-dispersed count data: Theory and applications. *Reliability: Theory & Applications*, 1(67), 33-51.
<https://doi.org/10.24412/1932-2321-2022-167-33-51>