



KKU SCIENCE JOURNAL

Journal Home Page : <https://ph01.tci-thaijo.org/index.php/KKUSciJ>

Published by the Faculty of Science, Khon Kaen University, Thailand



การประยุกต์ใช้ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม ในการพยากรณ์จำนวนอุบัติเหตุของรถบรรทุก จังหวัดสุรินทร์ Application of First-order Integer-valued Autoregressive Model for Forecasting Number of Truck Accidents in Surin Province, Thailand

กันติชา มณีกุล¹ กนกกรณ์ ลีโรจนาประภา^{1*} และ ชูเกียรติ พุดพรมราช²Kanticha Maneekul¹, Kanogkan Leerojanaprapa^{1*}, and Chookiat Pupprommarat²¹ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพมหานคร 10520²ภาควิชาวิทยาศาสตร์ประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา กรุงเทพมหานคร 10300¹Department of Statistics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand²Department of Applied Science, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, 10300, Thailand

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อพยากรณ์จำนวนอุบัติเหตุของรถบรรทุกรายเดือนที่เกิดขึ้นในจังหวัดสุรินทร์ ซึ่งข้อมูลการเกิดอุบัติเหตุเป็นข้อมูลจำนวนเต็มที่มีค่าตั้งแต่ 0 เป็นต้นไป จึงจัดเป็นข้อมูลแบบอนุกรมเวลาที่เป็จำนวนนับ (Time Series or Count Data) ที่มีรูปแบบแตกต่างจากข้อมูลอนุกรมเวลาทั่วไปที่มักเป็นจำนวนจริง และเมื่อข้อมูลมีตัวเลขที่เป็นศูนย์รวมอยู่ในชุดข้อมูลจำนวนมากจึงเลือกตัวแบบสำหรับข้อมูลที่มีค่าศูนย์เพื่อมาใช้ในการพยากรณ์ได้ ดังนั้นในการศึกษานี้จึงได้นำตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (INAR1) มาเปรียบเทียบกับตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองที่มีค่าศูนย์เพื่อ (ZIPINAR1) และตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบ (ZINBINAR(1)) โดยใช้เกณฑ์ค่าสนเทศของอะกะอิเกะ (AIC) ในการคัดเลือกตัวแบบ จากข้อมูลจำนวนอุบัติเหตุของรถบรรทุก ในจังหวัดสุรินทร์ ตั้งแต่เดือนมกราคม ปี พ.ศ. 2562 ถึงเดือนธันวาคม ปี พ.ศ. 2567 จำนวนทั้งหมด 72 ค่า มาสร้างตัวแบบการพยากรณ์ จากผลการศึกษาพบว่า ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ (ZIPINAR1) เหมาะสมที่สุดเมื่อเทียบกับตัวแบบอื่น ๆ โดยมีค่า AIC ต่ำที่สุดเท่ากับ 183.6693

ABSTRACT

This study aims to forecast the monthly number of truck-related accidents in Surin Province. The accident data are non-negative integers, as time series count data, which differ from conventional time series that are typically real numbers. In the presence of a large number of zeros, models suitable for zero-inflated data may be considered. Therefore, this study employed the First-order Integer-valued

*Corresponding Author, E-mail: kanogkan.le@kmitl.ac.th

Received date: 17 June 2025 | Revised date: 26 September 2025 | Accepted date: 29 October 2025

doi: 10.14456/kkuscij.2026.10

Autoregressive (INAR1) model and then compared it with the Zero-inflated Poisson INAR1 (ZIPINAR1) model and the Zero-inflated Negative Binomial INAR1 (ZINBINAR1) model. Model selection was based on the Akaike Information Criterion (AIC). The dataset comprised 72 monthly observations of truck accidents in Surin Province from January 2019 to December 2024. The results indicated that the ZIPINAR1 model provided the best fit, yielding the lowest AIC value of 183.6693 compared to the other models.

คำสำคัญ: จำนวนการเกิดอุบัติเหตุของรถบรรทุก ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม การแจกแจง-ปัวซองค่าศูนย์เพื่อ แจกแจงทวินามเชิงลบค่าศูนย์เพื่อ

Keywords: Number of Truck Accidents, First-order Integer-valued Autoregressive Model, Zero-inflated Poisson Distribution, Zero-inflated Negative Binomial Distribution

บทนำ

ระบบการขนส่งทางถนนเป็นระบบหลักของการคมนาคมในประเทศไทยมีประมาณทั้งสิ้น 4 28,426.2 ต้นโดยมีสัดส่วนสูงถึงร้อยละ 78.31 ของการขนส่งทั้งหมดในปี พ.ศ. 2565 (ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม, 2568) แม้ว่าการขนส่งทางถนนจะมีความสะดวกและสามารถรองรับการขนส่งสินค้าในปริมาณมากได้ แต่จำนวนรถบรรทุกที่เพิ่มขึ้นบนท้องถนนยังทำให้เกิดปัญหาการจราจรรุนแรงขึ้น และก่อให้เกิดความกังวลด้านความปลอดภัยของผู้ร่วมทาง อีกทั้งการขับรถในระยะทางไกลที่ทำให้พนักงานขับรถเกิดอาการอ่อนล้าและขาดสมาธิ ซึ่งเป็นการเพิ่มความเสี่ยงต่อการเกิดอุบัติเหตุบนท้องถนน (วุฒิ, 2562) งานวิจัยที่ผ่านมาได้ระบุว่า อุบัติเหตุที่เกี่ยวข้องกับรถบรรทุกมีปัจจัยสำคัญที่เกี่ยวข้องกับผู้ขับขี่ ยานพาหนะ สภาพถนน และสิ่งแวดล้อม (วรชัย, 2559) แนวโน้มดังกล่าวสะท้อนให้เห็นถึงความจำเป็นเร่งด่วนในการพัฒนากลยุทธ์เพื่อยกระดับความปลอดภัยในการขนส่งทางบก

จากสถิติข้อมูลอุบัติเหตุบนทางหลวงประจำปี พ.ศ. 2566 พบว่า อุบัติเหตุที่เกิดจากรถบรรทุกสับล้อขึ้นไปเกิดขึ้น 2,681 ครั้ง คิดเป็นร้อยละ 8.34 และอุบัติเหตุที่เกิดจากรถบรรทุกหกล้อขึ้นไปเกิดขึ้น 1,262 ครั้ง คิดเป็นร้อยละ 3.92 ซึ่งคิด 2 ใน 5 ของประเภทยานพาหนะที่เกิดอุบัติเหตุมากที่สุด (สำนักอำนวยการความปลอดภัย กรมทางหลวง, 2567) แม้ว่าจำนวนอุบัติเหตุที่เกิดจากรถบรรทุกจะไม่ได้มีสถิติการเกิดอุบัติเหตุเป็นอันดับที่หนึ่ง แต่อุบัติเหตุที่เกิดจากรถบรรทุกก่อให้เกิดความเสียหายรุนแรงกว่ายานพาหนะประเภทอื่น ๆ เนื่องจากตัวรถมีขนาดใหญ่และมีน้ำหนักมากกว่ารถยนต์ทั่วไปเป็นอย่างมาก ซึ่งเมื่อเกิดอุบัติเหตุ แรงปะทะที่เกิดจากการชนจะสูงกว่ามาก ทำให้เกิดความเสียหายและการบาดเจ็บที่รุนแรง อีกทั้งผลจากรถมีขนาดใหญ่อาจทำให้การจราจรบริเวณที่เกิดอุบัติเหตุติดขัด และหากสิ่งที่บรรทุกมาเป็นสินค้าจำพวกสารพิษ หรือน้ำมันเชื้อเพลิงจะส่งผลกระทบต่อสิ่งแวดล้อมเป็นอย่างมาก และจำเป็นต้องเตรียมพร้อมรับสถานการณ์ให้ทันทั่วทั้งหากเกิดอุบัติเหตุขึ้นในพื้นที่

จังหวัดสุรินทร์เป็นจังหวัดที่เป็นที่ตั้งเชิงยุทธศาสตร์ติดชายแดนประเทศกัมพูชา มีการส่งสินค้าผ่านช่องทางจุดผ่านแดนระหว่างประเทศที่ด่านช่องจอม อีกทั้งจังหวัดสุรินทร์เป็นจังหวัดที่มีการเชื่อมโยงกับจังหวัดอื่น ๆ เช่น จังหวัดบุรีรัมย์ ศรีสะเกษ และร้อยเอ็ด ซึ่งเป็นพื้นที่เกษตรกรรมและการผลิต ผ่านทางหลักหมายเลข 24 และถนนสายอื่น ๆ อีกทั้งยังเป็นจังหวัดที่เป็นศูนย์กลางทางโลจิสติกส์ กระจายสินค้าและวัตถุดิบ ดังนั้นการเตรียมพร้อมรับสถานการณ์จึงมีความจำเป็น และหากสามารถคาดการณ์จำนวนการเกิดอุบัติเหตุของรถบรรทุกจะทำให้จังหวัดสุรินทร์สามารถกำหนดนโยบายเชิงพื้นที่เพื่อจัดเตรียมเครื่องมือ อุปกรณ์ กำลังคน ที่ตรงตามความต้องการเฉพาะพื้นที่อย่างเพียงพอ เพื่อรับมือกับเหตุการณ์ดังกล่าวได้อย่างทันทั่วทั้ง โดยเฉพาะในช่วงเทศกาลที่มีการจราจรคับคั่ง

การพยากรณ์จำนวนครั้งการอุบัติเหตุที่เกิดขึ้นในแต่ละเดือนมีการศึกษาอย่างแพร่หลายในหลายประเทศ สำหรับตัวแบบการพยากรณ์ที่เลือกใช้ได้แก่ SARIMA (Seasonal Autoregressive Integrated Moving Average: SARIMA) และตัวแบบการปรับให้เรียบแบบเอกซ์โพเนนเชียล (Exponential Smoothing) โดย Rabbani *et al.* (2021) ได้รวบรวมงานวิจัยที่นำตัวแบบพยากรณ์ด้วย 2 วิธีนี้มาใช้ในการพยากรณ์การเกิดอุบัติเหตุ นอกจากนี้ยังมีงานวิจัยที่เลือกใช้แบบ ARIMA (Autoregressive Integrated Moving Average) เช่น การศึกษาของ Quddus (2008) Ramstedt (2008) Dadashova *et al.* (2016) Garcia-Ferrer *et al.* (2006) Zheng and Liu (2009) Sanusi *et al.* (2016) Parvareh *et al.* (2018) เป็นต้น ซึ่งส่วนใหญ่จะเป็นการใช้ข้อมูลจำนวนอุบัติเหตุที่รวมทุกประเภทรถยนต์ หรือจำนวนอุบัติเหตุรวมทั้งประเทศ ทำให้ข้อมูลที่น่าสนใจเป็นตัวเลขที่มีจำนวนมาก จนมีค่าใกล้เคียงกับข้อมูลที่เป็นจำนวนจริงทำให้สามารถใช้วิธีการพยากรณ์แบบแนวคิดแบบดั้งเดิม (classical approach) นำมาใช้ในการพยากรณ์ได้ผลเป็นอย่างดี

อย่างไรก็ตาม การพยากรณ์จำนวนครั้งการเกิดอุบัติเหตุที่เกิดขึ้นในระยะสั้น หรือเฉพาะประเภทรถยนต์ในพื้นที่รายจังหวัดนั้น มีประโยชน์ในการจัดการด้านความปลอดภัยทางการจราจรเชิงรุก ซึ่งสามารถติดตามสภาพการจราจรแบบเรียลไทม์ และสามารถดำเนินการแก้ปัญหาการจราจรได้อย่างทันท่วงที หากแต่มีข้อจำกัดเนื่องจากมีจำนวนครั้งการเกิดอุบัติเหตุไม่มาก หรือมีลักษณะเป็นตัวเลขแบบจำนวนนับ ซึ่งมีบางช่วงเวลาไม่มีการเกิดอุบัติเหตุ จึงมีข้อมูลที่มีค่าเป็นศูนย์จำนวนมาก หรืออาจเรียกว่ามีค่าศูนย์เฟ้อ (Cai *et al.*, 2024) การใช้วิธีดั้งเดิมอาจไม่เหมาะสมในการพยากรณ์ทั้งข้อมูลที่เป็นเลขศูนย์ หรือจำนวนนับที่ไม่เป็นศูนย์ (Feng, 2021) ดังนั้นจึงจำเป็นต้องเลือกวิธีการพยากรณ์การเกิดอุบัติเหตุที่เหมาะสมกับลักษณะของข้อมูลดังกล่าว ที่เรียกว่าการพยากรณ์ข้อมูลแบบอนุกรมเวลาที่เป็นจำนวนนับ (Time Series for Count Data) การกำหนดการแจกแจงแบบของข้อมูลจำนวนนับ เช่น การแจกแจงปัวซอง การแจกแจงทวินามเชิงลบ มาใช้แทนข้อสมมุติเรื่องการแจกแจงปกติ และยังการนับจำนวนอุบัติเหตุในช่วงเวลาสั้น ๆ ส่งผลให้โอกาสของการไม่พบอุบัติเหตุในช่วงดังกล่าว และหากเป็นกรณีเฉพาะที่มีข้อมูลในแต่ละช่วงเวลามีจำนวนครั้งการเกิดอุบัติเหตุเป็น 0 ครั้ง จำนวนมากกว่าปกติซึ่งเรียกว่าข้อมูลที่มีค่าศูนย์เฟ้อ

จากงานวิจัยที่เกี่ยวข้องในอดีต พบว่ามีการนำวิธีที่ใช้กับข้อมูลจำนวนนับที่มีค่าศูนย์เฟ้อ (Zero-inflated Count) กับข้อมูลความถี่ในการเกิดอุบัติเหตุรายช่วงถนน (Shankar *et al.*, 1997) จำนวนอุบัติเหตุบนถนนสายหลักในประเทศมาเลเซีย (Prasetijo and Musa, 2016) หรือทำนายจำนวนผู้เสียชีวิตจากอุบัติเหตุในประเทศไทย (Simmachan *et al.*, 2022) แต่เป็นการทำนายด้านเทคนิคสมการการถดถอยที่มีทั้งตัวแปรอิสระและตัวแปรตาม สำหรับงานวิจัยที่ใช้กับข้อมูลอนุกรมเวลาที่ผ่านมา พบว่า Brijs *et al.* (2008) แนะนำตัวแบบการถดถอยในตัวที่มีค่าเป็นจำนวนเต็ม (Integer Autoregressive) สำหรับข้อมูลอุบัติเหตุรายวัน นอกจากนี้ยังมีงานวิจัยของ Quddus (2008) ที่แนะนำตัวแบบ INAR/ZINAR มาใช้พยากรณ์อุบัติเหตุทางถนนในสหราชอาณาจักร หลังจากนั้น Barreto-Souza (2015) แนะนำ Zero-modified Geometric INAR(1) รองรับทั้งการแจกแจงค่าศูนย์เฟ้อ/ศูนย์น้อยกว่าปกติ และความแปรปรวน (variance) ของข้อมูลมากกว่า/น้อยกว่าค่าเฉลี่ย และสุดท้าย Garay *et al.* (2021) พบว่า ZI-INAR(1) ที่ใช้ขั้นตอนประมาณด้วย EM และบูตสเตรปมารวมในการกำหนดตัวแบบ อย่างไรก็ตาม จากการทบทวนวรรณกรรมพบว่าเทคนิคดังกล่าวยังไม่มีมีการนำมาประยุกต์ใช้กับการพยากรณ์ข้อมูลอนุกรมเวลาจำนวนอุบัติเหตุรถบรรทุกในประเทศไทย

ดังนั้นในงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพยากรณ์จำนวนครั้งการเกิดอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์ ด้วยการเปรียบเทียบตัวแบบทางสถิติ ทั้งที่เป็นการพยากรณ์ข้อมูลที่เป็นจำนวนเต็มแบบปกติ และเป็นจำนวนเต็มแบบที่มีค่าศูนย์เฟ้อมาประยุกต์ใช้ โดยเลือกเอาตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (First-order Integer-valued Autoregressive Model; INAR1) และตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มแบบมีค่าศูนย์เฟ้อ (First-order Integer-valued Autoregressive with Zero-inflated; ZINAR(1)) ด้วยการแจกแจงปัวซอง (Poisson Distribution)

และการแจกแจงทวินามเชิงลบ (Negative Binomial distribution) มาทำการเปรียบเทียบเพื่อหาตัวแบบการพยากรณ์ที่มีความเหมาะสมกับข้อมูลชุดนี้มากที่สุด

วิธีการดำเนินการวิจัย

การทำวิจัยครั้งนี้เป็นการศึกษาเกี่ยวกับการประยุกต์ใช้ตัวแบบการพยากรณ์อนุกรมเวลาสำหรับข้อมูลจำนวนนับสำหรับการพยากรณ์จำนวนอุบัติเหตุรายเดือนของรถบรรทุกในจังหวัดสุรินทร์ที่มีค่าศูนย์ในชุดข้อมูลเป็นจำนวนมาก ดังนั้นจึงได้เลือกเปรียบเทียบกับตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (First-order Integer-valued Autoregressive) ตามการแจกแจงแบบไม่ต่อเนื่องที่นิยมใช้สำหรับข้อมูลจำนวนนับซึ่งได้แก่ การแจกแจงปัวซอง การแจกแจงทวินามเชิงลบ มาเปรียบเทียบกับตัวแบบที่ใช้กับตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มค่าศูนย์เพื่อ (Zero-inflated) ด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ และการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ เพื่อประมาณค่าพารามิเตอร์ของทั้ง 4 ตัวแบบ เพื่อนำเสนอตัวแบบที่ดีที่สุด

จากการทบทวนวรรณกรรมพบว่า ตัวแบบการถดถอยในตัวที่นิยมนำมาใช้ในการวิเคราะห์ข้อมูลที่มีลักษณะเป็นจำนวนเต็มคือ ตัวแบบการถดถอยในตัวที่มีการแจกแจงปัวซอง แต่การแจกแจงปัวซองจะมีลักษณะเฉพาะที่สามารถอธิบายจำนวนครั้งของเหตุการณ์หรือสิ่งที่สนใจที่เกิดขึ้นในช่วงเวลาที่กำหนด โดยมีค่าเฉลี่ยเท่ากับค่าความแปรปรวน ซึ่งข้อมูลส่วนใหญ่จะมีค่าความแปรปรวนสูงกว่าค่าเฉลี่ย หรือเรียกว่าภาวะการกระจายเกินเกณฑ์ (overdispersion) ซึ่งอาจไม่เป็นไปตามทฤษฎีของตัวแบบการถดถอยในตัวที่มีการแจกแจงปัวซอง (Kusrisakul *et al.*, 2022) ดังนั้นในงานวิจัยนี้จึงมีการนำตัวแบบการถดถอยในตัวที่มีการแจกแจงทวินามเชิงลบ มาทำการเปรียบเทียบในกรณีที่เกิดปัญหาภาวะการกระจายเกินเกณฑ์ โดยมีทฤษฎีที่เกี่ยวข้องดังนี้

ตัวแบบการถดถอยในตัวที่มีค่าเป็นจำนวนเต็มสร้างขึ้นครั้งแรกโดย McKenzie (1985) และ Al-Osh and Alzaid (1987) และเหมาะกับข้อมูลอนุกรมเวลาที่มีค่าเป็นจำนวนนับที่ไม่มีค่าเป็นค่าลบ โดยตัวแบบที่มีการนำเสนอมีชื่อเรียกว่าตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (First-order Integer Autoregressive Model; INAR(1)) ซึ่งตัวแบบที่มีสหสัมพันธ์ในตัวอันดับที่หนึ่ง และได้กำหนดให้ค่าความคลาดเคลื่อนมีการแจกแจงปัวซอง (Poisson Distribution) จากนั้นตัวแบบได้มีการพัฒนาสู่การพยากรณ์กับข้อมูลจำนวนนับที่มีค่าเป็นศูนย์เป็นจำนวนมาก ที่เรียกว่า ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มค่าศูนย์เพื่อ (First-order Integer-valued Autoregressive with Zero-inflated; ZINAR(1)) โดยตัวแบบดังกล่าวได้ถูกนำมาพัฒนาต่อโดย Aghababaei *et al.* (2012) และ Garay *et al.* (2021) ซึ่งได้นำตัวแบบมาพัฒนาโดยใช้การแจกแจงของ Innovation ที่แตกต่างกัน เช่น นำมาใช้กับการแจกแจงแบบปัวซอง หรือแบบทวินามนิเสธ เป็นต้น และได้อธิบายตัวแบบที่เกี่ยวข้องดังต่อไปนี้

ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (First-order Integer-valued Autoregressive Model; INAR(1))

ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม ใช้ข้อมูลจำนวนเต็มจากช่วงเวลาก่อนหน้า (Autoregressive) 1 ค่า เพื่อพยากรณ์ค่าในเวลาปัจจุบัน

นิยามที่ 1 : กำหนดให้ X เป็นตัวแปรสุ่มที่มีค่าเป็นจำนวนเต็มไม่เป็นลบ เมื่อ $\alpha \in [0,1]$ และตัวดำเนินการที่ทำให้บางทวินาม (Binomial Thinning Operator) ' $\circ$ ' (Garay *et al.*, 2021) จะถูกกำหนดไว้ดังสมการที่ (1)

$$\alpha \circ X = \sum_{i=1}^X Z_i \quad (1)$$

เมื่อ Z_i เป็นลำดับของตัวแปรสุ่มแบบเบอร์นูลลี ที่มีการแจกแจงแบบอิสระและมีการแจกแจงเหมือนกัน (Independent Identically Distributed: i.i.d.) ของตัวแปรสุ่ม X ด้วย $P(Z_i = 1) = \alpha$

เมื่อพิจารณากระบวนการ INAR(1) ดังสมการที่ (2)

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z} \quad (2)$$

เมื่อ $\{X_t\}_{t \in \mathbb{Z}}$ เป็นลำดับของตัวแปรสุ่มที่มีค่าจำนวนเต็มที่ไม่เป็นลบ (Non-negative Integer-valued) หรือที่เรียกว่า innovation และ ε_t เป็นตัวแปรสุ่มความคลาดเคลื่อน โดย ε_t เป็นค่าที่เป็นอิสระจาก ε_{t-1} และ $\{X_t\}_{t \in \mathbb{Z}}$ สำหรับทุก ๆ ค่า t

กระบวนการ INAR(1) ตามคุณสมบัติความน่าจะเป็นแบบ Markov Chain (Weiss *et al.*, 2019) เป็นดังสมการที่ (3)

$$p_{x_i|x_{i-1}} : P(X_i = x_i | X_{i-1} = x_{i-1}) = \sum_{m=0}^{\min\{x_i, x_{i-1}\}} \binom{x_{i-1}}{m} \alpha^m (1-\alpha)^{x_{i-1}-m} P(\varepsilon_t = x_i - m) \quad (3)$$

โดยมีค่าเฉลี่ยและความแปรปรวนของ $\{X_t\}$ ดังนี้

$$E(\{X_t\}) = \frac{\mu_\varepsilon}{1-\alpha}$$

$$Var(\{X_t\}) = \frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2}$$

โดยมีการกำหนดฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไข (Conditional Likelihood Function) ดังสมการที่ (4) (Leonenko *et al.*, 2007)

$$L(\alpha, \theta | x) = \prod_{t=2}^T P(X_t = x_t | X_{t-1} = x_{t-1})$$

$$= \prod_{t=2}^T \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} P(\varepsilon_t = x_t - m) \quad (4)$$

โดยมีฟังก์ชันล็อกฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไข (Log Conditional Likelihood Function) ดังนี้

$$\ell(\alpha, \theta) = \sum_{t=2}^T \log \sum_{m=0}^{\min\{i, j\}} \binom{x_{i-1}}{m} \alpha^m (1-\alpha)^{x_{i-1}-m} P(\varepsilon_t = x_i - m) \quad (5)$$

สำหรับกระบวนการหาค่าประมาณพารามิเตอร์ของตัวแบบการถดถอยในตัวอันดับที่หนึ่งเมื่อกำหนดให้ ความคลาดเคลื่อน (ε_t) มีการแจกแจงที่แตกต่างกัน โดยการศึกษาครั้งนี้ได้ศึกษาทั้งหมด 4 การแจกแจง ได้แก่ การแจกแจงปัวซอง การแจกแจงทวินามเชิงลบ การแจกแจงปัวซองค่าศูนย์เพื่อ และการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ การหาค่าตัวประมาณพารามิเตอร์ของแต่ละตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม ดังนี้

1. ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง

ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (First-order Integer-valued Autoregressive with Poisson; PINAR(1)) ได้รับความนิยมอย่างมากในการศึกษาสำหรับข้อมูลจำนวนเต็มเพราะเป็นตัวแบบพื้นฐานที่ใช้การแจกแจงปัวซอง ซึ่งเป็นการแจกแจงเป็นที่รู้จักสำหรับข้อมูลจำนวนนับ เมื่อ $\{X_t\}$ เป็นไปตามกระบวนการ INAR(1) ด้วยการแจกแจงปัวซอง โดยมีตัวดำเนินการที่ทำให้บางทวินาม และ ε_t มีการแจกแจงปัวซอง (Poisson Distribution) ด้วยพารามิเตอร์ μ ดังนั้น $\{X_t\}_{t \geq 1}$ ซึ่งเรียกว่า ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง หรือ PINAR(1) โดยใช้ตัวดำเนินการที่ทำให้บางทวินาม และ ε_t เป็นตัวแปรสุ่มความคลาดเคลื่อนที่มีการแจก

แจกแจงพร้อมทั้ง $\alpha \circ X_{t-1}$ และ ε_t ซึ่งเป็นอิสระต่อกัน เมื่อกำหนดให้ $\varepsilon_t \sim \text{Poisson}(\mu)$ โดยฟังก์ชันมวลความน่าจะเป็นของตัวแปรสุ่ม ε_t ดังสมการที่ (6)

$$P(\varepsilon_t = v) = \frac{e^{-\mu} \mu^v}{v!}, \quad v=0,1,2,\dots \quad (6)$$

ซึ่งมีค่าเฉลี่ยและความแปรปรวนดังนี้

$$\begin{aligned} E(\varepsilon_t) &= (1-\rho)\mu \\ \text{Var}(\varepsilon_t) &= (1-\rho)\mu(1+\rho\mu) \end{aligned}$$

ได้ฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขดังนี้

$$P(X_i = x_i | X_{i-1} = x_{i-1}) = \sum_{m=0}^{\min\{x_i, x_{i-1}\}} \binom{x_{i-1}}{m} \alpha^m (1-\alpha)^{x_{i-1}-m} \frac{\mu^{x_i-m} e^{-\mu}}{(x_i-m)!} \quad (7)$$

และฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไข และฟังก์ชันล็อกภาวะน่าจะเป็นแบบมีเงื่อนไขของตัวแบบ PINAR(1) แสดงดังสมการที่ (8) และ (9) ตามลำดับ

$$L(\alpha, \mu | \{X_t\}) = \frac{\mu^{x_1} e^{-\mu}}{x_1!} \prod_{t=2}^T \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} \frac{\mu^{x_t-m} e^{-\mu}}{(x_t-m)!} \quad (8)$$

$$\ell(\alpha, \mu) = \log \left(\frac{\mu^{x_1} e^{-\mu}}{x_1!} \right) + \sum_{t=2}^T \log \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} \frac{\mu^{x_t-m} e^{-\mu}}{(x_t-m)!} \quad (9)$$

สำหรับการประมาณพารามิเตอร์ตัวแบบ PINAR(1) ด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบมีเงื่อนไข จะนำ $\ell(\alpha, \mu)$ มาหาอนุพันธ์อันดับที่ 1 เทียบกับ α และ μ พร้อมทั้งกำหนดให้เท่ากับศูนย์ หลังจากนั้นแก้ระบบสมการไม่เป็นเชิงเส้น

$U(\alpha, \mu) = \left(\frac{\partial \ell(\alpha, \mu)}{\partial \alpha}, \frac{\partial \ell(\alpha, \mu)}{\partial \mu} \right) = 0$ โดยใช้วิธีเทคนิคการทำซ้ำ (Iterative Technique) ด้วยอัลกอริทึมนิวตันราฟสัน (Newton-Raphson Algorithm)

2. ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบ

ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบ (First-order Integer-valued Autoregressive with Negative Binomial: NBINAR(1)) มีการศึกษามาโดย Al-Osh and Aly (1992) โดยเป็นแบบทางเลือกหนึ่งของปัญหาการเกิดการกระจายเกินเกณฑ์ ในการแจกแจงปัวซอง ดังนั้นจึงมีการศึกษาการแจกแจงทวินามเชิงลบเป็นการแจกแจงทางเลือกหนึ่งสำหรับแก้ปัญหาการเกิดการกระจายเกินเกณฑ์ (สุชีราภรณ์และจิราพรรณ, 2564) โดย $\{X_t\}$ เป็นไปตามกระบวนการ INAR(1) ด้วยการแจกแจงทวินามเชิงลบ โดยใช้ตัวดำเนินการที่ทำให้บางทวินาม และ ε_t มีการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ μ และ ϕ พร้อมทั้ง $\alpha \circ X_{t-1}$ และ ε_t ซึ่งเป็นอิสระต่อกัน โดยฟังก์ชันมวลความน่าจะเป็นของตัวแปรสุ่ม ε_t ดังสมการที่ (10)

$$P(\varepsilon_t = v) = \frac{\Gamma(\phi + v)}{\Gamma(v+1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^v \left(\frac{\phi}{\mu + \phi} \right)^\phi, \quad v=0,1,2,\dots \quad (10)$$

เมื่อ $\mu \geq 0$ และ $\phi > 0$ เป็น Dispersion พารามิเตอร์ และ $\Gamma(\cdot)$ เป็น ฟังก์ชันแกมมา ซึ่งมีค่าเฉลี่ยและความแปรปรวนดังนี้

$$E(\varepsilon_t) = \mu$$

$$\text{Var}(\varepsilon_t) = \mu \left(1 + \frac{\mu}{\phi} \right)$$

ได้ฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขดังนี้

$$P(X_i = x_i | X_{i-1} = x_{i-1}) = \sum_{m=0}^{\min\{x_i, x_{i-1}\}} \binom{x_{i-1}}{m} \alpha^m (1-\alpha)^{x_{i-1}-m} \frac{\Gamma(x_i - k + \phi)}{\Gamma(x_i - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_i - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \quad (11)$$

โดยมีฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไข และฟังก์ชันล็อกภาวะน่าจะเป็นแบบมีเงื่อนไขของตัวแบบ NBINAR(1) แสดงดังสมการที่ (12) และ (13) ตามลำดับ

$$\begin{aligned} L(\alpha, \mu, \phi | \{X_t\}) &= \frac{\Gamma(x_1 - k + \phi)}{\Gamma(x_1 - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_1 - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \prod_{t=2}^T \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} \\ &\times \frac{\Gamma(x_t - k + \phi)}{\Gamma(x_t - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_t - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \end{aligned} \quad (12)$$

$$\begin{aligned} \ell(\alpha, \mu, \phi) &= \log \left[\frac{\Gamma(x_1 - k + \phi)}{\Gamma(x_1 - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_1 - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \right] + \sum_{t=2}^T \log \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} \\ &\times \frac{\Gamma(x_t - k + \phi)}{\Gamma(x_t - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_t - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \end{aligned} \quad (13)$$

สำหรับการประมาณพารามิเตอร์ตัวแบบ NBINAR(1) ด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบมีเงื่อนไข โดยการนำ $\ell(\alpha, \mu, \phi)$ มาหาอนุพันธ์อันดับที่ 1 เทียบกับ α , μ และ ϕ พร้อมทั้งกำหนดให้เท่ากับศูนย์ หลังจากนั้นแก้ระบบสมการไม่เป็นเชิงเส้น $U(\alpha, \mu, \phi) = \left(\frac{\partial \ell(\alpha, \mu, \phi)}{\partial \alpha}, \frac{\partial \ell(\alpha, \mu, \phi)}{\partial \mu}, \frac{\partial \ell(\alpha, \mu, \phi)}{\partial \phi} \right) = 0$ โดยใช้วิธีเทคนิคการทำซ้ำ (Iterative Technique) ด้วยอัลกอริทึมนิวตันราฟสัน (Newton-Raphson Algorithm)

3. ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ

ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ (First-order Integer-valued Autoregressive with Zero-inflated Poisson; ZIPINAR(1)) เป็นตัวแบบทางสถิติที่พัฒนาจากตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม ใช้ในการจัดการกับข้อมูลที่เป็นจำนวนเต็ม และมีค่าเป็นศูนย์ (Zero-inflated Count Data) โดยใช้การแจกแจงปัวซองค่าศูนย์เพื่อ โดย Garay *et al.* (2021) ได้นำเสนอการนำกระบวนการ ZINAR(1) เมื่อ $\{X_t\}$ เป็นไปตามกระบวนการ INAR(1) ด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ (ZIPINAR(1)) โดยใช้ตัวดำเนินการที่ทำให้บางทวินาม และ ε_t เป็นตัวแปรสุ่มที่มีการแจกแจงปัวซองค่าศูนย์เพื่อ โดยมีพารามิเตอร์ μ และ ρ พร้อมทั้ง $\alpha \circ X_{t-1}$ และ ε_t เป็นอิสระต่อกัน โดยฟังก์ชันมวลความน่าจะเป็นของตัวแปรสุ่ม ε_t ดังสมการที่ (14)

$$P(\varepsilon_t = v) = \begin{cases} \rho + (1-\rho)e^{-\mu} & , v=0 \\ (1-\rho) \frac{e^{-\mu} \mu^v}{v!} & , v \geq 1 \end{cases} \quad (14)$$

ซึ่งมีค่าเฉลี่ยและความแปรปรวนดังนี้

$$E(\varepsilon_t) = (1-\rho)\mu$$

$$Var(\varepsilon_t) = (1-\rho)\mu(1+\rho\mu)$$

ฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขดังนี้

$$P(X_i = x_i | X_{i-1} = x_{i-1}) = \sum_{m=0}^{\min\{x_i, x_{i-1}\}} \binom{x_{i-1}}{m} \alpha^m (1-\alpha)^{x_{i-1}-m} [\rho + (1-\rho)e^{-\mu}]^{\delta_t} \left[(1-\rho) \frac{\mu^{x_i-m} e^{-\mu}}{(x_i-m)!} \right]^{1-\delta_t} \quad (15)$$

$$\text{โดยที่} \quad \delta_t = \begin{cases} 1, & x_t - m = 0 \\ 0, & x_t - m \neq 0 \end{cases}$$

โดยมีฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไข และฟังก์ชันล็อกความน่าจะเป็นแบบมีเงื่อนไขของตัวแบบ ZIPINAR(1) แสดงดังสมการที่ (15) และ (16) ตามลำดับ

$$L(\alpha, \mu, \rho | \{X_t\}) = \prod_{t=2}^T \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} [\rho + (1-\rho)e^{-\mu}]^{\delta_t} \left[(1-\rho) \frac{\mu^{x_t-m} e^{-\mu}}{(x_t-m)!} \right]^{1-\delta_t} \quad (16)$$

$$\ell(\alpha, \mu, \rho) = \sum_{t=2}^T \log \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1-\alpha)^{x_{t-1}-m} [\rho + (1-\rho)e^{-\mu}]^{\delta_t} \left[(1-\rho) \frac{\mu^{x_t-m} e^{-\mu}}{(x_t-m)!} \right]^{1-\delta_t} \quad (17)$$

สำหรับการประมาณพารามิเตอร์ตัวแบบ ZIPINAR(1) ด้วยวิธีความน่าจะเป็นสูงสุดแบบมีเงื่อนไข จะนำ $\ell(\alpha, \mu, \rho)$ มาหาอนุพันธ์อันดับที่ 1 เทียบกับ α μ และ ρ พร้อมทั้งกำหนดให้เท่ากับศูนย์ หลังจากนั้นแก้ระบบสมการไม่เป็นเชิงเส้น $U(\alpha, \mu, \rho) = \left(\frac{\partial \ell(\alpha, \mu, \rho)}{\partial \alpha}, \frac{\partial \ell(\alpha, \mu, \rho)}{\partial \mu}, \frac{\partial \ell(\alpha, \mu, \rho)}{\partial \rho} \right) = 0$ โดยใช้วิธีเทคนิคการทำซ้ำ (iterative technique) ด้วย อัลกอริทึมนิวตันราฟสัน (Newton-Raphson Algorithm)

4. ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ

ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ (First-order Integer-valued Autoregressive with Zero-inflated Negative Binomial; ZINBINAR(1)) เป็นตัวแบบทางสถิติที่นำเสนอวิธีการพยากรณ์ค่าเป็นจำนวนเต็ม โดยใช้การแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ เมื่อ $\{x_t\}$ เป็นไปตามกระบวนการ INAR(1) ด้วยการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ (ZINBINAR(1)) โดยใช้ตัวดำเนินการที่ทำให้บางทวินาม และ ε_t เป็นตัวแปรสุ่มความคลาดเคลื่อนที่มีการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อที่มีพารามิเตอร์ ρ μ และ ϕ พร้อมทั้ง $\alpha \circ x_{t-1}$ และ ε_t เป็นอิสระต่อกัน (Weiss et al., 2019) โดยฟังก์ชันมวลความน่าจะเป็นของตัวแปรสุ่ม ε_t ดังสมการที่ (18)

$$P(\varepsilon_t = v) = \begin{cases} \rho + (1-\rho) \left(\frac{\phi}{\mu + \phi} \right)^\phi, & v=0 \\ (1-\rho) \frac{\Gamma(\phi+v)}{\Gamma(v+1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^v \left(\frac{\phi}{\mu + \phi} \right)^\phi, & v \geq 1 \end{cases} \quad (18)$$

เมื่อ $\mu \geq 0$ และ $\phi > 0$ เป็น Dispersion พารามิเตอร์ และ $\Gamma(\cdot)$ เป็น ฟังก์ชันแกมมา

ซึ่งมีค่าเฉลี่ยและความแปรปรวนดังนี้

$$E(\varepsilon_t) = (1 - \rho)\mu$$

$$\text{Var}(\varepsilon_t) = (1 - \rho)\mu \left(1 + \frac{\mu}{\phi} + \rho\mu \right)$$

ฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขดังนี้

$$P(X_i = x_i | X_{i-1} = x_{i-1}) = \sum_{m=0}^{\min\{x_i, x_{i-1}\}} \binom{x_{i-1}}{m} \alpha^m (1 - \alpha)^{x_{i-1}-m} \left[\rho + (1 - \rho) \left(\frac{\phi}{\mu + \phi} \right)^\phi \right]^{\delta_t} \times \left[(1 - \rho) \frac{\Gamma(x_t - k + \phi)}{\Gamma(x_t - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_t - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \right]^{1 - \delta_t} \quad (19)$$

โดยที่

$$\delta_t = \begin{cases} 1, & x_t - m = 0 \\ 0, & x_t - m \neq 0 \end{cases}$$

โดยมีฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไข และฟังก์ชันล็อกความน่าจะเป็นแบบมีเงื่อนไขของตัวแบบ ZINBINAR(1) แสดงดังสมการที่ (19) และ (20) ตามลำดับ

$$L(\alpha, \mu, \phi, \rho | \{X_t\}) = \prod_{t=2}^T \sum_{m=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{m} \alpha^m (1 - \alpha)^{x_{t-1}-m} \left[\rho + (1 - \rho) \left(\frac{\phi}{\mu + \phi} \right)^\phi \right]^{\delta_t} \times \left[(1 - \rho) \frac{\Gamma(x_t - k + \phi)}{\Gamma(x_t - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_t - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \right]^{1 - \delta_t} \quad (20)$$

$$\ell(\alpha, \mu, \phi, \rho) = \sum_{t=2}^T \log \sum_{m=0}^{\min\{i, j\}} \binom{x_{i-1}}{m} \alpha^m (1 - \alpha)^{x_{i-1}-m} \times \left[\rho + (1 - \rho) \left(\frac{\phi}{\mu + \phi} \right)^\phi \right]^{\delta_t} \times \left[(1 - \rho) \frac{\Gamma(x_t - k + \phi)}{\Gamma(x_t - k + 1)\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi} \right)^{x_t - k} \left(\frac{\phi}{\mu + \phi} \right)^\phi \right]^{1 - \delta_t} \quad (21)$$

สำหรับการประมาณพารามิเตอร์จากตัวแบบ ZINBINAR(1) ด้วยวิธีความน่าจะเป็นสูงสุดแบบมีเงื่อนไข โดยการนำ $\ell(\alpha, \mu, \phi, \rho)$ มาหาอนุพันธ์อันดับที่ 1 เทียบกับ α μ ϕ และ ρ พร้อมทั้งกำหนดให้เท่ากับศูนย์ หลังจากนั้นแก้ระบบสมการไม่เป็นเชิงเส้น $U(\alpha, \mu, \phi, \rho) = 0$

$$U(\alpha, \mu, \phi, \rho) = \left(\frac{\partial \ell(\alpha, \mu, \phi, \rho)}{\partial \alpha}, \frac{\partial \ell(\alpha, \mu, \phi, \rho)}{\partial \mu}, \frac{\partial \ell(\alpha, \mu, \phi, \rho)}{\partial \phi}, \frac{\partial \ell(\alpha, \mu, \phi, \rho)}{\partial \rho} \right) = 0$$

โดยใช้วิธีเทคนิคการทำซ้ำ (Iterative Technique) ด้วยอัลกอริทึมนิวตันราฟสัน (Newton-Raphson Algorithm)

เกณฑ์การคัดเลือกตัวแบบ

ในงานวิจัยนี้จะใช้เกณฑ์การคัดเลือกตัวแบบโดยเกณฑ์สารสนเทศของอะกะอิเกะ (AIC) ในการเปรียบเทียบตัวแบบที่เหมาะสมที่สุด โดยพิจารณาจากฟังก์ชันล็อกภาวะน่าจะเป็นสูงสุด โดยเกณฑ์สารสนเทศของอะกะอิเกะ (AIC) มีสูตรดังสมการที่ (22)

$$AIC = -2(LL) + 2k \quad (22)$$

เมื่อ LL คือ ค่าฟังก์ชันล็อกภาวะน่าจะเป็นสูงสุด (maximum log likelihood function) จากการประมาณค่าพารามิเตอร์ และ k คือ จำนวนพารามิเตอร์ของตัวแบบ ซึ่งตัวแบบที่ให้ค่า AIC ต่ำสุดจะเป็นตัวแบบที่เหมาะสมที่สุด

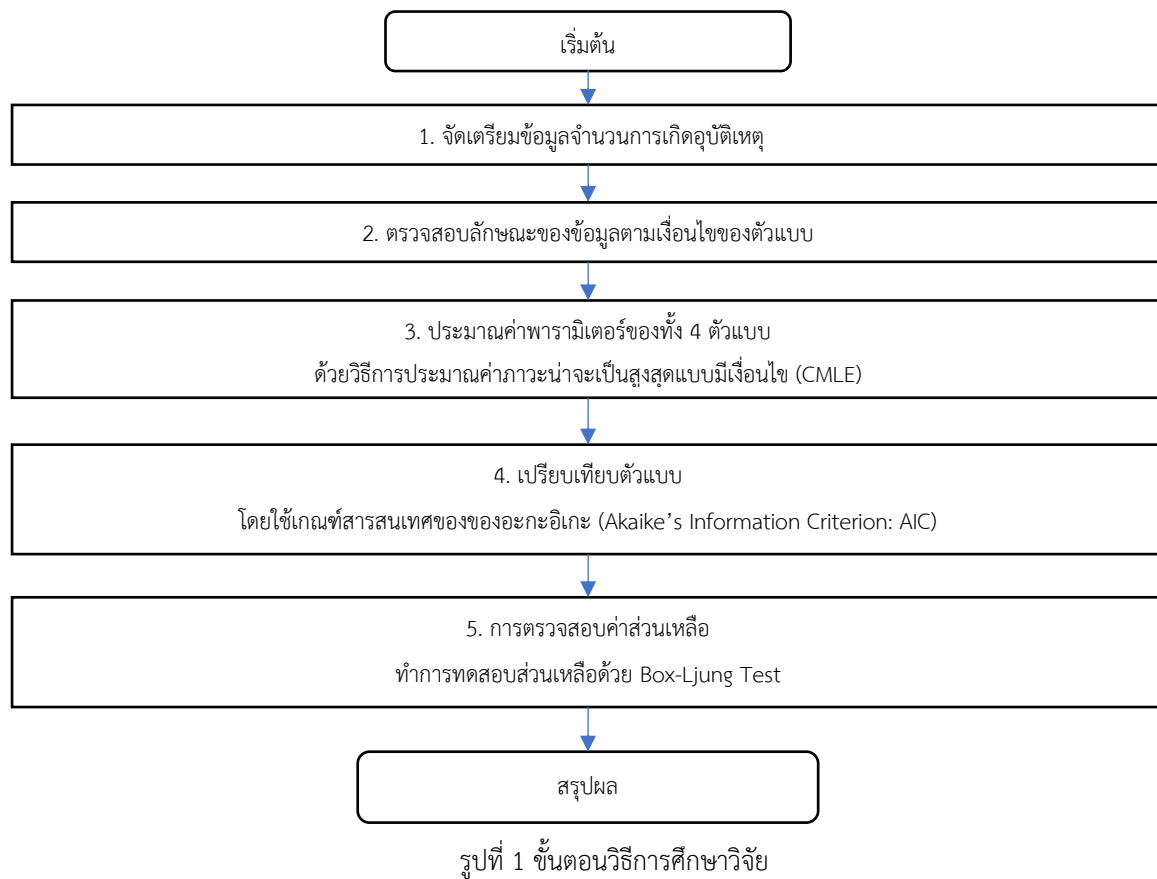
งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการเลือกใช้วิธีการพยากรณ์จำนวนการเกิดอุบัติเหตุ พบว่ามีการพยากรณ์ด้วยตัวแบบสมการการถดถอยแบบทั่วไปและการถดถอยในตัว เช่น Orozco *et al.* (2021) ซึ่งพบว่า Quddus (2008) ได้ทำการเปรียบเทียบการพยากรณ์ข้อมูลแบบจำนวนนับ ด้วยตัวแบบการถดถอยในตัวที่มีค่าเป็นจำนวนเต็ม (INAR) เทียบกับ ตัวแบบ Autoregressive Integrated Moving Average (ARIMA) และได้แบ่งกลุ่มการวิเคราะห์เป็น 2 กลุ่มคือ จำนวนอุบัติเหตุที่มีตัวเลขมากจากการรวมจำนวนอุบัติเหตุระดับพื้นที่ขนาดใหญ่หรือจำนวนครั้งต่อระยะเวลานานเช่นต่อ 1 ปี หรือ การพิจารณาขนาดจำนวนอุบัติเหตุต่อพื้นที่ย่อย ๆ หรือจำนวนอุบัติเหตุต่อเดือน โดยทำการเปรียบเทียบตัวแบบการถดถอยในตัวที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (INARP) และ ARIMA โดยพบว่า ตัวแบบตัวแบบการถดถอยในตัวที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (INARP) สามารถพยากรณ์ได้ดีกว่า ARIMA เมื่อเป็นกลุ่มที่ 2 ข้อมูลตัวเลขน้อย เป็นแบบการพิจารณาขนาดจำนวนอุบัติเหตุต่อพื้นที่ย่อย ๆ

นอกจากการใช้วิธี ARIMA ยังมีการประยุกต์ใช้การถดถอยในตัวที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (INARP) มีการศึกษาในด้านต่าง ๆ เช่น Freeland and McCabe (2004) ศึกษาการใช้ตัวแบบ Poisson Autoregressive Time-series Model สำหรับข้อมูลจำนวนนับที่ประยุกต์ใช้กับข้อมูลจำนวนการเคลมเอาประกันรายเดือนเนื่องจากการบาดเจ็บระหว่างทำงานซึ่งเป็นข้อมูลจาก Workers Compensation Board (WCB) แห่งบริติชโคลัมเบีย ประเทศแคนาดา ในขณะที่ Orozco *et al.* (2021) นำเสนอแบบจำลองการถดถอยในตัวอันดับที่ 1 สำหรับตัวแบบผสมจำนวนเต็ม ที่มีการใช้ปัวซอง (First-order Integer-valued Autoregressive Process with Poisson Innovations) เรียกว่า POMINAR(1) โดยผสมกับตัวดำเนินการที่ทำให้บางแบบทวินามและแบบปัวซอง โดยนำมาประยุกต์ใช้กับการทำนายจำนวนคดีด้านการลักทรัพย์ของสถานีตำรวจในเมือง Pittsburgh นอกจากนี้ Lord and Mannering (2010) นำเสนอการทบทวนวรรณกรรมที่เกี่ยวข้องกับข้อมูลความถี่ในการเฉี่ยวชน และการวิเคราะห์จุดแข็งจุดอ่อนของวิธีที่มีการนำมาศึกษาในเรื่องนี้ ในงานวิจัยได้รวม Zero-inflated Poisson and Negative Binomial เป็น 1 ใน 16 วิธีที่รวบรวมมาโดยระบุว่าจุดแข็งในการพยากรณ์มีค่าสังเกตจำนวนอุบัติเหตุ เป็น 0 จำนวนมาก ได้ดีโดยเน้นตัวแบบการถดถอย มากกว่าตัวแบบอนุกรมเวลา

วิธีการศึกษา

ในการกำหนดตัวแบบที่เหมาะสมสำหรับการพยากรณ์จำนวนครั้งการเกิดอุบัติเหตุที่มีรูปแบบของข้อมูลแตกต่างจากข้อมูลอนุกรมเวลาทั่วไปที่สามารถใช้วิธีดั้งเดิม ดังนั้นการวิจัยนี้จะนำเสนอขั้นตอนการกำหนดตัวแบบที่เหมาะสมสำหรับการพยากรณ์ดังรูปที่ 1 โดยมีรายละเอียดดังต่อไปนี้



1. ข้อมูลที่ใช้ในการศึกษา

การศึกษาครั้งนี้ใช้ข้อมูลทุติยภูมิ โดยได้รับการอนุเคราะห์ข้อมูลจากจำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์ ซึ่งเป็นข้อมูลที่ได้รับการอนุเคราะห์จาก กลุ่มงานแผนความปลอดภัย สำนักสวัสดิภาพการขนส่งทางบก กรมการขนส่งทางบก (2567) โดยทำการเก็บรวบรวมข้อมูลอุบัติเหตุรายเดือนตั้งแต่เดือนมกราคม ปี พ.ศ. 2562 ถึงเดือนธันวาคม ปี พ.ศ. 2567 จำนวนทั้งสิ้น 72 เดือน และทำการตรวจสอบคุณภาพข้อมูลเบื้องต้น ไม่พบข้อมูลศูนย์หาย หรือข้อมูลผิดปกติ จากนั้นนั้นข้อมูลทั้ง 72 ค่า แสดงตารางที่ 1

ตารางที่ 1 ข้อมูลอนุกรมเวลาของจำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์ตั้งแต่เดือนมกราคม ปี พ.ศ. 2562 ถึงเดือนธันวาคม ปี พ.ศ. 2567

ปี	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.
2562	2	0	0	0	0	0	1	1	1	0	0	1
2563	1	0	0	0	1	0	0	0	0	0	0	0
2564	3	0	0	0	0	0	0	1	1	0	1	0
2565	3	2	2	0	3	0	2	2	1	0	1	1
2566	3	2	4	6	1	1	1	0	0	0	0	0
2567	2	1	2	1	0	2	0	1	3	2	2	0

งานวิจัยที่ใช้ INAR (Integer-valued Autoregressive Model) ส่วนใหญ่ไม่ได้เน้นการพยากรณ์ข้อมูลใหม่ในระยะยาว และงานวิจัยนี้สนใจการประมาณพารามิเตอร์ของตัวแบบในกระบวนการอนุกรมเวลาที่มีข้อมูลเป็นจำนวนเต็มและพยากรณ์ในระยะสั้นจึงไม่ได้แบ่งข้อมูลชุดฝึกฝน และชุดทดสอบ

2. ตรวจสอบข้อสมมุติของตัวแบบ

ก่อนการหาตัวแบบ INAR ผู้วิจัยต้องทำการตรวจสอบข้อมูลก่อนว่ามีค่าเป็นจำนวนเต็ม ได้ทำการตรวจสอบข้อมูลต้องมีลักษณะที่อยู่กับที่ในสถานะนิ่ง (stationary) และตรวจสอบข้อมูลว่าเป็นข้อมูลที่มีรูปแบบของความแปรปรวนเกินเกณฑ์ หรือมีลักษณะที่ความแปรปรวนใกล้เคียงกับค่าเฉลี่ยหรือไม่ เพื่อประกอบเหตุผลของการเลือกกลุ่มตัวแบบทั้ง 4 วิธีมาทำการเปรียบเทียบ

3. การประมาณค่าพารามิเตอร์

สำหรับประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม จะประมาณค่าด้วยวิธีภาวน่าจะเป็นสูงสุดแบบมีเงื่อนไข (CMLE) โดยมี 4 ตัวแบบดังนี้

- 1) ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (PINAR1)
- 2) ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบ (NBINAR1)
- 3) ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ (ZIPINAR1)
- 4) ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ (ZINBINAR1)

4. เกณฑ์การคัดเลือกตัวแบบ

การเปรียบเทียบประสิทธิภาพของตัวแบบทั้ง 4 ตัวแบบ เปรียบเทียบโดยใช้เกณฑ์สารสนเทศของอะไคเกะ (Akaike's Information Criterion: AIC) ในการคัดเลือกตัวแบบ ซึ่งตัวแบบที่ให้ค่า AIC ต่ำสุดจะเป็นตัวแบบที่เหมาะสมกับข้อมูลมากที่สุด นอกจากนี้ยังเลือกใช้การทดสอบ Likelihood-ratio สำหรับ Nested Models ระหว่าง PINAR(1) และ ZIPINAR(1) เพื่อยืนยันความเหมาะสมของ ZIPINAR(1) อีกด้วย

5. การตรวจสอบค่าส่วนเหลือ

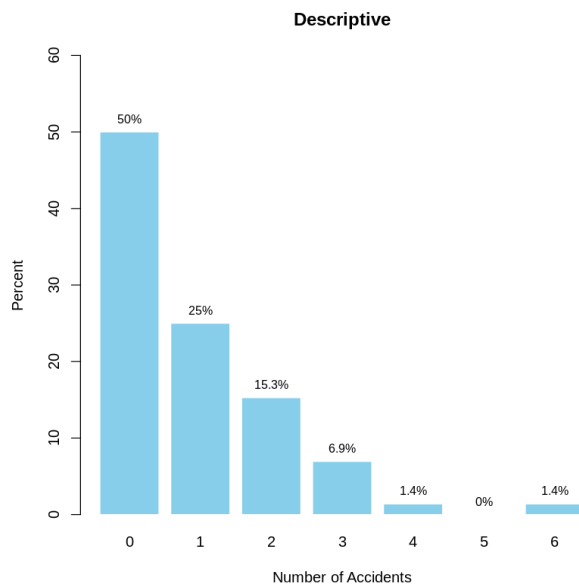
เศษเหลือคือความแตกต่างระหว่างค่าจริงและค่าพยากรณ์จากตัวแบบที่กำหนด การวิเคราะห์เศษเหลือเป็นกระบวนการสุดท้ายเพื่อให้มั่นใจได้ว่าตัวแบบมีความเหมาะสมหรือไม่ หากตัวแบบเหมาะสมกับข้อมูลจะพบว่า ค่าเศษเหลือจะมีคุณสมบัติการเป็นเชิงสุ่ม (white noise) นั่นคือเศษเหลือจะไม่ขึ้นกับเวลา โดยมีค่าเฉลี่ยเป็นศูนย์ ความแปรปรวนคงที่ และไม่มีสหสัมพันธ์ในตัว ในขั้นตอนนี้จะทำการทดสอบด้วย Box-Ljung Test เพื่อยืนยันว่าตัวแบบที่เลือกมานั้นเหมาะสมหรือไม่

ผลการวิจัยและวิจารณ์ผล

ผลลัพธ์ที่ได้จากการศึกษาสามารถสรุปผลในแต่ละขั้นตอนได้ดังต่อไปนี้

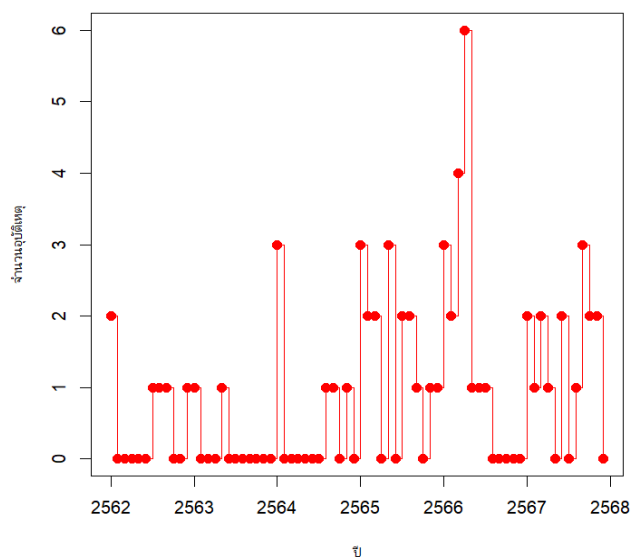
1. การนำเสนอข้อมูลเบื้องต้น

จากการนำข้อมูลอนุกรมเวลาของจำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์รายเดือน ตั้งแต่เดือนมกราคม ปี พ.ศ. 2562 ถึงเดือนธันวาคม ปี พ.ศ. 2567 จำนวน 72 ค่าสังเกต แสดงดังตารางที่ 1 พบว่าข้อมูลจำนวนครั้งของการเกิดอุบัติเหตุดังกล่าวเป็นข้อมูลจำนวนนับที่มีค่าอยู่ระหว่าง 0 - 6 ครั้งต่อเดือน โดยมีจำนวนครั้งการเกิดอุบัติเหตุสูงสุดในเดือนเมษายน พ.ศ. 2566 และมีค่าเฉลี่ย เท่ากับ 0.90 ค่าความแปรปรวน เท่ากับ 1.41 และค่า และค่าภาวะการกระจายเกินเกณฑ์ เท่ากับ 1.57 เมื่อนำมาวิเคราะห์สถิติเบื้องต้นแสดงดังรูปที่ 2



รูปที่ 2 ร้อยละจำนวนเดือนที่มีจำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์

พบว่า ข้อมูลจำนวนอุบัติเหตุมีลักษณะการกระจายข้อมูลแบบเบ้ขวา และมีค่ามีค่าอยู่ระหว่าง 0 - 6 ครั้ง นอกจากนี้ยังพบว่าข้อมูลไม่มีทั้งองค์ประกอบของแนวโน้มและฤดูกาล โดยแสดงดังรูปที่ 3



รูปที่ 3 จำนวนอุบัติเหตุที่เกิดขึ้นในแต่ละเดือนที่ระหว่าง มกราคม พ.ศ. 2562 ถึง ธันวาคม พ.ศ. 2568

2. การตรวจสอบสถานะนิ่ง Stationarity

ขั้นตอนนี้เป็นการตรวจสอบข้อมูลให้มีลักษณะที่อยู่นิ่ง (stationary) นำเสนอการทดสอบ 2 วิธี Augmented Dickey-Fuller Test และ Phillips-Perron Unit Root test แสดงผลการทดสอบสมมติฐานดังตารางที่ 2 และมีการกำหนดสมมติฐานดังต่อไปนี้

สมมติฐาน H_0 : ข้อมูลไม่อยู่ในสถานะนิ่ง (non-stationary)

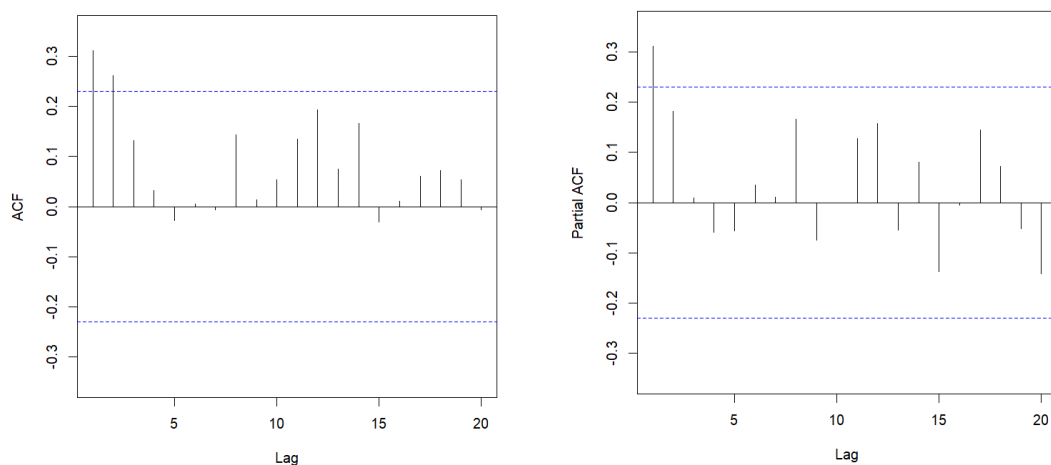
H_1 : ข้อมูลอยู่ในสถานะนิ่ง (stationary)

ผลการทดสอบสมมติฐานด้วยวิธี Augmented Dickey-Fuller Test พบว่ามีค่า p -value = 0.02738 วิธี Phillips - Perron Unit Root test มีค่า p -value = 0.01 ซึ่งพบน้อยกว่า $\alpha = 0.05$ จึงปฏิเสธ H_0 สรุปได้ว่าข้อมูลอยู่ในสภาวะนิ่ง (stationary) ที่ระดับนัยสำคัญ 0.05

ตารางที่ 2 การทดสอบข้อมูลที่มีลักษณะที่อยู่ในสภาวะนิ่ง (stationary)

การทดสอบ	สถิติทดสอบ (Test Statistic)	ค่าพี (p -value)	สรุป
Augmented Dickey-Fuller Test	- 3.7429	0.02738*	ข้อมูลอยู่ในสภาวะนิ่ง
Phillips-Perron Unit Root test	- 59.324	0.01*	ข้อมูลอยู่ในสภาวะนิ่ง

* ที่ระดับนัยสำคัญ 0.05



รูปที่ 4 ลักษณะการเคลื่อนไหวของข้อมูล ค่า ACF และ PACF

จากนั้นทำการวิเคราะห์ฟังก์ชันออโตคอร์เรเลชัน (Autocorrelation Function: ACF) ในรูปที่ 4 (ซ้าย) และ ฟังก์ชันออโตคอร์เรเลชันบางส่วน (Partial Autocorrelation Function: PACF) ในรูปที่ 4 (ขวา) พบว่าข้อมูลมีลักษณะที่อยู่ในสภาวะนิ่ง โดยมีกราฟ ACF มีลักษณะลดลงอย่างค่อยเป็นค่อยไป ขณะที่กราฟ PACF มีค่าสูงอย่างมีนัยสำคัญที่เวลาหน่วง (lag) ลำดับที่ 1 และลดลงเข้าสู่ศูนย์อย่างรวดเร็วในลำดับถัดไป ลักษณะดังกล่าวบ่งชี้ว่าข้อมูลจำนวนอุบัติเหตุในแต่ละเดือนมีความสัมพันธ์กับค่าก่อนหน้าเพียงหนึ่งช่วงเวลา ซึ่งสอดคล้องกับสมมติฐานของตัวแบบการถดถอยในตัวเองอันดับที่หนึ่ง (First Order Autoregressive Model: AR(1)) ดังนั้นจึงสามารถใช้ตัวแบบ AR(1) ในการอธิบายและพยากรณ์แนวโน้มของข้อมูลอุบัติเหตุได้อย่างเหมาะสมกับข้อมูลจำนวนเต็มกับข้อมูลชุดนี้ จากนั้นนำมาสร้างตัวแบบการถดถอยในตัวเองอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (INAR1) ที่มีรูปแบบการแจกแจงปัวซอง การแจกแจงทวินามเชิงลบ และตัวแบบการถดถอยในตัวเองอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มค่าศูนย์เพื่อ (ZINAR(1)) ที่มีรูปแบบการแจกแจงปัวซองค่าศูนย์เพื่อ และการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ ด้วยวิธีตัวประมาณค่าภาวะน่าจะเป็นสูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood Estimators) ด้วยโปรแกรม R แสดงผลการประมาณค่าพารามิเตอร์ในแต่ละตัวแบบ และค่า AIC ในลำดับถัดไป

3. การประมาณค่าพารามิเตอร์ของแต่ละตัวแบบ

จากตารางที่ 2 จะเห็นว่าตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง ค่าศูนย์เพื่อ (ZIPINAR(1)) และตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงทวินามเชิงลบ ค่าศูนย์เพื่อ (ZINBINAR(1)) จะมีค่าประมาณพารามิเตอร์ ซึ่งคือค่าของสัดส่วนการเกิดค่าศูนย์ และแทนด้วยสัญลักษณ์ ซึ่งพบว่าพารามิเตอร์นี้จะไม่พบในตัวแบบที่ไม่ใช่ชนิดที่มีค่าศูนย์เพื่อ ซึ่งได้แก่ การถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (PINAR(1)) และทวินามเชิงลบ (NBINAR(1)) ในขณะที่ค่าประมาณพารามิเตอร์ และแทนด้วยสัญลักษณ์ มีเฉพาะในตัวแบบการถดถอยที่เกี่ยวข้องกับการแจกแจงทวินามเชิงลบเท่านั้น โดยค่าพารามิเตอร์ ซึ่งใช้วัดระดับภาวะการกระจายเกินเกณฑ์ ของการแจกแจงทวินามเชิงลบ พบว่า NBINAR(1) มีค่าอยู่ที่ร้อยละ 8.84 คือความแปรปรวนของข้อมูลมากกว่าค่าเฉลี่ยเพียงเล็กน้อย ส่วนของ ZINBINAR(1) มีค่าอยู่ที่ร้อยละ 36.74 คือความแปรปรวนของข้อมูลมากกว่าค่าเฉลี่ยอย่างชัดเจน

ตารางที่ 3 ค่า AIC และค่าประมาณพารามิเตอร์ของตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง การแจกแจงทวินามเชิงลบ การแจกแจงปัวซองค่าศูนย์เพื่อ และการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ

ตัวแบบ	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\mu}$	$\hat{\phi}$	AIC
1. PINAR(1)	0.242	NA ⁽¹⁾	0.666	NA ⁽²⁾	188.9853
2. NBINAR(1)	0.262	NA ⁽¹⁾	0.647	0.884	184.8891
3. ZIPINAR(1)	0.296	0.496	1.224	NA ⁽²⁾	183.6693
4. ZINBINAR(1)	0.294	0.485	1.201	36.74	185.7017

Note: ⁽¹⁾ ไม่ใช่พารามิเตอร์ของตัวแบบทวินามนิเสธ (Negative Binomial)

⁽²⁾ ไม่ใช่พารามิเตอร์ของตัวแบบปัวซอง (Poisson)

ผลเปรียบเทียบตัวแบบทั้ง 4 ตัวแบบ โดยใช้เกณฑ์การคัดเลือกตัวแบบด้วยเกณฑ์สารสนเทศของอะกะอิเกะ (AIC) แสดงดังตารางที่ 3 เมื่อพบว่า ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ (ZIPINAR(1)) เป็นตัวแบบพยากรณ์จำนวนอุบัติเหตุน้อยของรถบรรทุกที่เกิดในจังหวัดสุรินทร์ที่เหมาะสมที่สุด เนื่องจากให้ค่า AIC ต่ำที่สุด (AIC เท่ากับ 183.6693) เมื่อเทียบกับตัวแบบอื่น ๆ

โดยค่าประมาณพารามิเตอร์ ρ ของ ZIPINAR(1) มีค่าสูงกว่า ZINBINAR(1) เล็กน้อย ซึ่งหมายความว่ามีการเกิดค่าศูนย์อยู่ที่ร้อยละ 49.6 และ 48.5 ตามลำดับ แสดงให้เห็นว่าทั้งสองตัวแบบสามารถอธิบายการเกิดค่าศูนย์ในข้อมูลได้ใกล้เคียงกัน แต่ ZIPINAR(1) คาดการณ์สัดส่วนค่าศูนย์ได้สูงกว่าเล็กน้อย กล่าวคือ ในตัวแบบ ZIPINAR(1) เกือบครึ่งหนึ่งของจำนวนเดือนทั้งหมดมีโอกาสที่จะไม่เกิดอุบัติเหตุของรถบรรทุก และค่าประมาณพารามิเตอร์ $\alpha = 0.294$ หมายถึงความน่าจะเป็นที่จำนวนอุบัติเหตุเดือนก่อนหน้า ($t - 1$) จะอธิบายความแปรปรวนเดือนปัจจุบันได้ประมาณร้อยละ 29 และมีค่าเฉลี่ยการเกิดอุบัติเหตุอยู่ที่ 1.224 ครั้งต่อเดือน

แทนความน่าจะเป็นที่หน่วยการนับในช่วงเวลา $t - 1$ จะส่งผลต่อค่าการนับในช่วงเวลา t ผ่านกลไก binomial thinning

ลำดับต่อไปเป็นการเปรียบเทียบตัวแบบ ZIPINAR(1) ที่มีพารามิเตอร์เพิ่มขึ้นจากตัวแบบ PINAR(1) ที่เป็นตัวแบบที่ซ้อนกัน (Nested) ทำให้ตัวแบบนี้ดีขึ้นหรือไม่ จึงทำการทดสอบด้วย Likelihood Ratio Test โดยการกำหนดสมมติฐานดังนี้

สมมติฐาน H_0 : PINAR(1) เหมาะสม

H_1 : ZIPINAR(1) เหมาะสมมากกว่า

ผลการทดสอบสมมติฐานด้วยวิธี Likelihood Ratio Test ดังตารางที่ 4 พบว่ามีค่า Likelihood Ratio Test = 7.31 และ $p\text{-value} = 0.0068$ ซึ่งพบว่าน้อยกว่า $\alpha = 0.05$ จึงปฏิเสธ H_0 สรุปได้ว่า พารามิเตอร์ ρ ที่เพิ่มขึ้นมาจากตัวแบบ ZIPINAR(1) สามารถเพิ่มความสามารถพยากรณ์ได้ดีมากกว่า PINAR(1) อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05

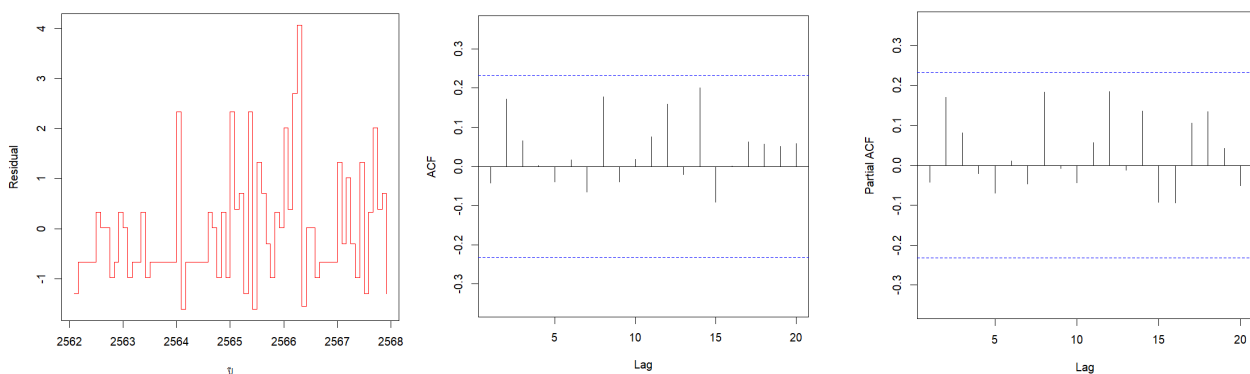
ตารางที่ 4 การทดสอบ Likelihood Ratio Test

การทดสอบ	สถิติทดสอบ (Test Statistic)	องศาความเป็นอิสระ (Degree of Freedom)	ค่าพี ($p\text{-value}$)
Likelihood Ratio Test	7.3161	1	0.0068*

* ที่ระดับนัยสำคัญ 0.05

4. การตรวจสอบค่าส่วนเหลือ

ลักษณะการเคลื่อนไหวของค่าส่วนเหลือ (Residual) จากตัวแบบ ZIPINAR(1) ที่คัดเลือกกว่าเป็นตัวแบบที่เหมาะสมที่สุดจากการเปรียบเทียบ 4 ตัวแบบจากหัวข้อก่อนหน้านี้ โดยนำค่าส่วนเหลือมาทำการทดสอบและสร้างกราฟ ACF และ PACF แสดงผลดังรูปที่ 5 จากกราฟแสดงความสัมพันธ์ระหว่างค่าเศษเหลือและเวลา พบว่าค่าเศษเหลือไม่มีแนวโน้ม มีการกระจายอย่างสุ่มโดยมีค่าเศษเหลือกระจายอยู่รอบค่าศูนย์ และมีความแปรปรวนค่อนข้างคงที่ นอกจากนี้ยังมีการทดสอบค่าเศษเหลือว่ามีสหสัมพันธ์ในตัวเอง (Autocorrelation) หรือไม่ด้วยการทดสอบ Ljung-Box



รูปที่ 5 ลักษณะการเคลื่อนไหวของค่าส่วนเหลือ (Residual) จากตัวแบบ ZIPINAR(1) และค่า ACF และ PACF

การทดสอบเศษเหลือโดยใช้ Ljung-Box แสดงผลดังตารางที่ 5 ได้ผลดังนี้

สมมติฐาน H_0 : ไม่มีสหสัมพันธ์ในตัวเอง (Autocorrelation) ของเศษเหลือ ในช่วง lag ที่ทดสอบ (เศษเหลือเป็น White Noise)

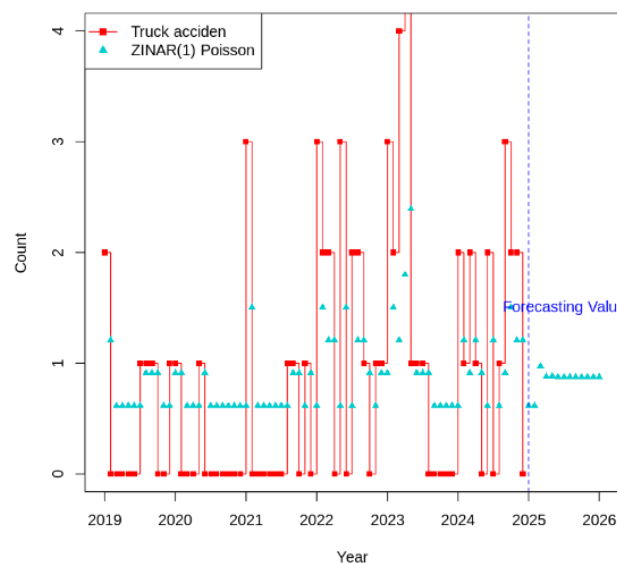
H_1 : มีอย่างน้อยหนึ่งค่ามีสหสัมพันธ์ในตัวเอง (Autocorrelation) ของเศษเหลือ ในช่วง lag ที่ทดสอบ (เศษเหลือไม่ใช่ White Noise)

ตารางที่ 5 การทดสอบเศษเหลือโดยใช้ Ljung-Box

การทดสอบ	สถิติทดสอบ (Test Statistic)	องศาความเป็นอิสระ (Degree of Freedom)	ค่าพี ($p\text{-value}$)
Ljung-Box	5.9328	10	0.8209

จากผลการทดสอบ Ljung-Box ของเศษเหลือ พบว่า มีค่า $p\text{-value} = 0.8209 > 0.05$ จึงยอมรับ H_0 สรุปได้ว่าเศษเหลือไม่มีของเศษเหลือ (Autocorrelation) ที่มีนัยสำคัญที่ 0.05 ในช่วง lag ที่ทดสอบ (เศษเหลือเป็น White Noise) ซึ่งผลการทดสอบสอดคล้องกับกราฟที่แสดงค่า ACF และ PACF ของเศษเหลือ ที่มีค่าอยู่ภายในขอบแสดงว่าไม่มีหลักฐานเพียงพอที่จะบอกว่ามีสหสัมพันธ์ในตัว (Autocorrelation) ของเศษเหลือ ดังนั้นจึงสามารถสรุปได้ว่าตัวแบบ ZIPINAR(1) มีความเหมาะสม

โดยผู้วิจัยได้นำตัวแบบ ZIPINAR(1) ที่เหมาะสมที่สุดไปพยากรณ์จำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์เทียบกับค่าจริง ในเดือนต่าง ๆ แสดงการเปรียบเทียบค่าสังเกตและค่าพยากรณ์ในแต่ละเดือนดังรูปที่ 6 และแสดงค่าพยากรณ์ล่วงหน้าเดือนต่าง ๆ ในปี พ.ศ. 2568 (ค.ศ 2025) แสดงดังตารางที่ 6



รูปที่ 6 การเปรียบเทียบค่าจริงและค่าพยากรณ์จำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์จากตัวแบบ ZIPINAR(1)

ตารางที่ 6 ค่าพยากรณ์จำนวนการเกิดอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์รายเดือนของปี พ.ศ. 2568 จากตัวแบบ ZIPINAR(1)

เดือน	ค่าพยากรณ์
มกราคม	0.6168960
กุมภาพันธ์	0.7994972
มีนาคม	0.8535472
เมษายน	0.8695460
พฤษภาคม	0.8742816
มิถุนายน	0.8756834
กรกฎาคม	0.8760983
สิงหาคม	0.8762211
กันยายน	0.8762574
ตุลาคม	0.8762682
พฤศจิกายน	0.8762714
ธันวาคม	0.8762723

สรุปผลและอภิปรายผลการวิจัย

ในงานวิจัยนี้นำเสนอการประยุกต์ใช้ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม (INAR1) จำนวน 4 ตัวแบบ ประกอบด้วย ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง (PINAR1) การแจกแจงทวินามเชิงลบ (NBINAR1) การแจกแจงปัวซองค่าศูนย์เพื่อ (ZIPINAR1) และการแจกแจงทวินามเชิงลบค่าศูนย์เพื่อ (ZINBINAR1) มาประยุกต์ใช้กับชุดข้อมูลจำนวนอุบัติเหตุของรถบรรทุกรายเดือนในจังหวัดสุรินทร์ ตั้งแต่เดือนมกราคม พ.ศ. 2562 ถึงเดือนธันวาคม พ.ศ. 2567 จำนวน 72 ค่า โดยทำการประมาณค่าพารามิเตอร์ และเปรียบเทียบตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็ม เพื่อนำเสนอตัวแบบที่เหมาะสมที่สุด ที่ได้จากการเปรียบเทียบประสิทธิภาพของตัวแบบทั้ง 4 ตัวแบบ ด้วยเกณฑ์การคัดเลือกตัวแบบโดยเกณฑ์สารสนเทศของอะกะอิเกะ (AIC) ผลการศึกษาพบว่า ตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ (ZIPINAR1) ให้ค่าพยากรณ์แม่นยำที่สุด และผ่านเกณฑ์การตรวจสอบตัวแบบ โดยประมาณค่าสัดส่วนการเกิดค่าศูนย์ (ρ) อยู่ที่ร้อยละ 49.6 หรือเกือบครึ่งหนึ่งของจำนวนเดือนทั้งหมดโอกาสที่จะไม่เกิดอุบัติเหตุของรถบรรทุกเกิดขึ้น และ $\alpha = 0.294$ หมายถึง ความน่าจะเป็นที่จำนวนอุบัติเหตุเดือนก่อนหน้าสามารถอธิบายความแปรปรวนเดือนปัจจุบันได้ประมาณร้อยละ 29

เนื่องจากข้อมูลชุดนี้จะมีค่าศูนย์มากกว่าปกติ อยู่ที่ประมาณร้อยละ 50 ของข้อมูลทั้งหมด ดังนั้นตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มค่าศูนย์เพื่อให้ค่าพยากรณ์ที่แม่นยำกว่า ซึ่งมีความสอดคล้องกับทฤษฎีการแจกแจงค่าศูนย์เพื่อ สำหรับการวิเคราะห์ข้อมูลในแต่ละครั้ง ควรคำนึงถึงการเลือกตัวแบบที่นำมาวิเคราะห์ให้เหมาะสมกับข้อมูล จากการศึกษาครั้งนี้ให้ผลสอดคล้องกับงานวิจัยของ Jazia *et al.* (2012) ที่ได้ทำการศึกษาตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง และตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ โดยได้นำมาใช้เพื่อจำลองข้อมูลจำนวนเหตุการณ์ในช่วงเวลาที่ต่อเนื่องกัน แต่เนื่องจากข้อมูลมีภาวะการกระจายเกินเกณฑ์ ข้อมูลมีค่าศูนย์เพื่อ ผลการศึกษางานวิจัยของ Jazia *et al.* (2012) จึงได้ผลว่าตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองค่าศูนย์เพื่อ เหมาะสมกับชุดข้อมูลที่มีค่าศูนย์เพื่อมากกว่าตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซอง อย่างไรก็ตามในการศึกษาครั้งต่อไปอาจพิจารณนำตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มค่าศูนย์เพื่อด้วยการแจกแจงรูปแบบอื่น ๆ เช่น อินเวอร์สเกาส์เซียนค่าศูนย์เพื่อ มาทำการเปรียบเทียบ เพื่อเปรียบเทียบกับผลการศึกษาของ Garay *et al.* (2021) ที่พบว่าตัวแบบที่เหมาะสมที่สุด คือตัวแบบการถดถอยในตัวอันดับที่หนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองอินเวอร์สเกาส์เซียนค่าศูนย์เพื่อ (ZIPIGINAR(1))

ในการศึกษาครั้งนี้เป็นการใช้ข้อมูลที่มีขนาดตัวอย่างขนาดเล็ก (72 ค่า) และไม่ได้นำตัวปัจจัยอื่น ๆ มาทำการวิเคราะห์ร่วมด้วย เช่น สภาพอากาศ การก่อสร้างถนน รวมถึงการแยกประเภทของรถบรรทุกตามน้ำหนัก ดังนั้นในการศึกษาครั้งต่อไป ควรจะมีการนำปัจจัยที่เกี่ยวข้องเข้ามาพิจารณาร่วมด้วย เพื่อจะสามารถพัฒนาตัวแบบโดยการนำการแจกแจงไม่ต่อเนื่องอื่น ๆ ที่มีความเหมาะสมกับข้อมูลเข้ามาร่วมในการศึกษาเพิ่มเติมได้ เช่น การแจกแจงเรขาคณิต (Geometric Distribution) การแจกแจงปัวซอง-อินเวอร์สเกาส์เซียน (Poisson-Inverse Gaussian Distribution) การแจกแจงปัวซอง-อินเวอร์สเกาส์เซียนค่าศูนย์เพื่อ (Zero-inflated Poisson-Inverse Gaussian Distribution) หรือสามารถขยายตัวแบบ INAR ใช้ร่วมกับ GLM ทดลองเปรียบเทียบตัวแบบที่ได้ในกลุ่มข้อมูลอนุกรมเวลาแบบจำนวนนับ (Time-series for Count Data) กับวิธีการเรียนรู้ของเครื่อง (machine learning) เช่น SVM หรือ LSTM นอกจากนี้อาจเป็นการศึกษาวิเคราะห์ข้อมูลจำนวนอุบัติเหตุรถบรรทุกเปรียบเทียบเชิงพื้นที่ กับจังหวัดข้างเคียง เป็นต้น หรือนำมาประยุกต์ใช้กับข้อมูลอนุกรมเวลาที่เป็นจำนวนเต็มและมีค่าศูนย์เพื่อ หรือข้อมูลอื่น ๆ ที่คาดว่าจะเหมาะสมกับตัวแบบอื่น ๆ ต่อไป

เอกสารอ้างอิง

- วรชัย บุญฤทธิผล. (2559). การวิเคราะห์ปัจจัยเชิงสาเหตุที่มีอิทธิพลต่อการเกิดอุบัติเหตุของพนักงานขับรถบรรทุก. วิทยานิพนธ์หลักสูตรวิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยบูรพา. ชลบุรี. 132 หน้า.
- วุฒิ พลบูรณ์. (2562). คู่มือนักสืบอุบัติเหตุโดยสารขนาดใหญ่. ศูนย์วิชาการเพื่อความปลอดภัยทางถนน (ศวปถ.). มหาวิทยาลัยเทคโนโลยีสุรนารี. นครราชสีมา. 76 หน้า.
- ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม. (2568). ปริมาณการขนส่งสินค้าภายในประเทศ. แหล่งข้อมูล: <https://datagov.mot.go.th/dataset/freight-dom>. ค้นเมื่อวันที่ 17 กันยายน 2568.
- สำนักอำนวยการความปลอดภัย กรมทางหลวง. (2567). จำนวนอุบัติเหตุของรถบรรทุกที่เกิดในจังหวัดสุรินทร์. แหล่งข้อมูล: <http://bhs.doh.go.th/download/accident>. ค้นเมื่อวันที่ 27 เมษายน 2567.
- สุชีราภรณ์ บุญยธีร และ จิราพรรณ สุนทรโชติ (2564). ตัวแบบการถดถอยในตัวอันดับหนึ่งที่มีค่าเป็นจำนวนเต็มด้วยการแจกแจงปัวซองลินด์เลย์สองพารามิเตอร์นัยทั่วไปบนฐานของตัวดำเนินการทำให้บางทวินามลบ. วารสารคณิตศาสตร์. 66(704): 63 - 77.
- Aghababaei Jazi, M., Jones, G. and Lai, C.D. (2022). Integer valued AR (1) with geometric innovations. *Journal of the Iranian Statistical Society* 11(2): 173 - 190.
- Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis* 8(3): 261 - 275. doi: 10.1111/j.1467-9892.1987.tb00438.x.
- Al-Osh, M.A. and Aly, E.E.A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics-Theory and Methods* 21(9): 2483 - 2492. doi: 10.1080/03610929208830925.
- Barreto-Souza, W. (2015). Zero-modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. *Journal of Time Series Analysis* 36(6): 839 - 852. doi: 10.1111/jtsa.12131.
- Brijs, T., Karlis, D. and Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis & Prevention* 40(3): 1180 - 1190. doi: 10.1016/j.aap.2008.01.001.
- Cai, B., Quddus, M., Wang, X. and Miao, Y. (2024). New modeling approach for predicting disaggregated time-series traffic crashes. *Transportation research board* 2678(3): 637 - 648. doi: 10.1177/03611981231182703
- Dadashova, B., Arenas-Ramírez, B., Mira-McWilliams, J. and Aparicio-Izquierdo, F. (2016). Methodological development for selection of significant predictors explaining fatal road accidents. *Accident Analysis & Prevention* 90: 82 - 94.
- Feng, C.X. (2021). A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of statistical distributions and applications* 8(1): 8. doi: 10.1186/s40488-021-00121-4.
- Freeland, R.K. and McCabe, B.P. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of time series analysis* 25(5): 701 - 722.

- Garay, A.M., Medina, F.L., Jales I.C.S. and Bertail, P. (2021). First-order integer valued AR processes with zero-inflated innovations. In: *Nonstationary Systems: Theory and Applications: Contributions to the 13th Workshop on Nonstationary Systems and Their Applications*, February 3 - 5, 2020, Grodek nad Dunajcem, Poland. Cham: Springer. 19 - 40.
- Jazia, M.A., Jones, G. and Lai, C.D. (2012). First-order integer valued AR processes with zero-inflated Poisson innovations. *Journal of Time Series Analysis* 33(6): 954 - 963.
- Leonenko, N.N., Savani, V. and Zhigljavsky, A.A. (2007). Autoregressive negative binomial processes. In *Annales de l'ISUP* 51(1 - 2): 25 - 47.
- Lord, D. and Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation research part A: policy and practice* 44(5): 291 - 305.
- McKenzie, E. (1985). Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association* 21(4): 645 - 650. doi: 10.1111/j.1752-1688.1985.tb05379.x.
- Orozco, D.L.R., Sales, L.O.F., Fernández, L.M.Z. and Pinho, A.L.S. (2021). A new mixed first-order integer-valued autoregressive process with Poisson innovations. *ASTA Advances in Statistical Analysis* 105(4): 559 - 580.
- Parvareh, M., Karimi, A., Rezaei, S., Woldemichael, A., Nili, S., Nouri, B. and Nasab, N.E. (2018). Assessment and prediction of road accident injuries trend using time-series models in Kurdistan. *Burns & trauma* 6: 9. doi: 10.1186/s41038-018-0111-6.
- Prasetijo, J. and Musa, W.Z. (2016). Modeling zero-inflated regression of road accidents at Johor federal road F001. In *MATEC Web of Conferences* 47: 03001. doi: 10.1051/matecconf/20164703001.
- Quddus, M. A. (2008). Time series count data models: an empirical application to traffic accidents. *Accident analysis & prevention* 40(5): 1732 - 1741.
- Rabbani, M.B.A., Musarat, M.A., Alaloul, W.S., Rabbani, M.S., Maqsoom, A., Ayub, S., Bukhari, H. and Altaf, M. (2021). A comparison between seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing (ES) based on time series model for forecasting road accidents. *Arabian Journal for Science and Engineering* 46(11): 11113 - 11138.
- Ramstedt, M. (2008). Alcohol and fatal accidents in the United States-a time series analysis for 1950-2002. *Accident Analysis & Prevention* 40(4): 1273 - 1281. doi: 10.1016/j.aap.2008.01.008.
- Sanusi, R.A., Adebola, F.B., and Adegoke, N.A. (2016). Cases of road traffic accident in Nigeria: a time series approach. *Mediterranean journal of social sciences* 7(2): 542 - 552.
- Shankar, V., Milton, J. and Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis & Prevention* 29(6): 829 - 837. doi: 10.1016/S0001-4575(97)00052-3.
- Simmachan, T., Wongsai, N., Wongsai, S. and Lersuwansri, R. (2022). Modeling road accident fatalities with underdispersion and zero-inflated counts. *PLOS ONE* 17(11): e0269022. doi: 10.1371/journal.pone.0269022.

- Weiss, C.H., Homburg, A. and Puig, P. (2019). Testing for zero inflation and overdispersion in INAR (1) models. *Statistical Papers* 60: 823 - 848.
- Zheng, X. and Liu, M. (2009). An overview of accident forecasting methodologies. *Journal of Loss Prevention in the process Industries* 22(4): 484 - 491. doi: 10.1016/j.jlp.2009.03.005.

