

ตัวแบบพยากรณ์ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขตของประเทศไทย

Forecasting Model for Pneumonia Cases in 13 Health Districts in Thailand

วฐา มินเสน^{1*} และ ปรารถนา มินเสน²

Watha Minsan^{1*} and Pradthana Minsan²

¹ ศูนย์วิจัยวิทยาการข้อมูล ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ เชียงใหม่ ประเทศไทย

² ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่ เชียงใหม่ ประเทศไทย

¹Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

²Department of Mathematics and Statistics, Faculty of Science and Technology, Chiang Mai Rajabhat University, Chiang Mai, Thailand

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาตัวแบบพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขตของประเทศไทย จากข้อมูลทุติยภูมิ Health Data Center กระทรวงสาธารณสุข ด้วยวิธีการวิเคราะห์อนุกรมเวลา 4 วิธี คือ วิธีแยกส่วนประกอบ วิธีทำให้เรียบ วิธีบอซ-เจนกินส์ และวิธีพยากรณ์รวม โดยข้อมูลจำแนกเป็นรายเดือนอยู่ในช่วงเดือนมกราคม พ.ศ. 2558 ถึง เดือนธันวาคม พ.ศ. 2565 ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพแต่ละเขตใช้จำนวนข้อมูลที่แตกต่างกันตามความสมบูรณ์ของข้อมูล แบ่งข้อมูลออกเป็น 2 ชุด คือชุดข้อมูลฝึกฝนเพื่อสร้างตัวแบบพยากรณ์ และชุดข้อมูลทดสอบเพื่อตรวจสอบความแม่นยำของตัวแบบพยากรณ์ โดยใช้เกณฑ์วัดความแม่นยำด้วยร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบสมมาตร (Symmetric Mean Absolute Percentage Error: sMAPE) ผลการวิจัยเพื่อหาตัวแบบที่เหมาะสมพบว่า เขตสุขภาพที่ 1 ถึง 6 และ 8 9 10 12 และ 13 มีตัวแบบพยากรณ์ที่เหมาะสมคือ ตัวแบบพยากรณ์รวมโดยใช้วิธีถ่วงน้ำหนักด้วยสัมประสิทธิ์การถดถอย และตัวแบบพยากรณ์ที่เหมาะสมของเขตสุขภาพที่ 7 คือ วิธีแยกส่วนประกอบที่มีส่วนประกอบของแนวโน้มและฤดูกาลรูปแบบคูณและเขตสุขภาพที่ 11 ตัวแบบพยากรณ์ที่เหมาะสมคือ ตัวแบบพยากรณ์บอซ-เจนกินส์

คำสำคัญ: ผู้ป่วยโรคปอดอักเสบ แยกส่วนประกอบ การทำให้เรียบ บอซ-เจนกินส์ การพยากรณ์รวม

ABSTRACT

The objective of this research is to evaluate a proposed forecasting model for predicting pneumonia cases across all 13 health districts in Thailand. Monthly time series data from January 2015 to December 2022, obtained from the Health Data Center of the Ministry of Public Health, were utilized. Each health district utilized a varying number of data points depending on data completeness. The data were categorized into two sets: a training dataset for model construction and a test dataset for assessing model accuracy. The forecasting models were evaluated using the symmetric mean absolute percentage error (sMAPE). The results revealed that the combined forecasting approach using regression analysis demonstrated advantages over other methods for health districts 1 to 6, 8, 9, 10, 12, and 13. For health district 7, the most appropriate model was the multiplicative decomposition model with trend and seasonal components. Similarly, for health district 11, the Box-Jenkins model was found to be the most suitable.

KEYWORDS: pneumonia patient, decomposition, smoothing, Box-Jenkins, combined forecasting

*Corresponding Author: wathaminsan@gmail.com

Received: 13/06/2023; Revised: 25/01/2024; Accepted: 16/02/2024

1. บทนำ

โรคปอดอักเสบ หรือปอดบวม (Pneumonia) เป็นโรคติดต่อทางเดินหายใจที่พบเห็นกันมากทั่วไป จากผลการ

สำรวจและการจัดลำดับ 10 โรคที่มีอัตราป่วยของผู้ป่วยในต่อประชากรหนึ่งแสนคน (รวมทุกการวินิจฉัยโรค) พ.ศ. 2558 พบว่า โรคปอดอักเสบสูงเป็นอันดับที่ 7 ของกลุ่ม

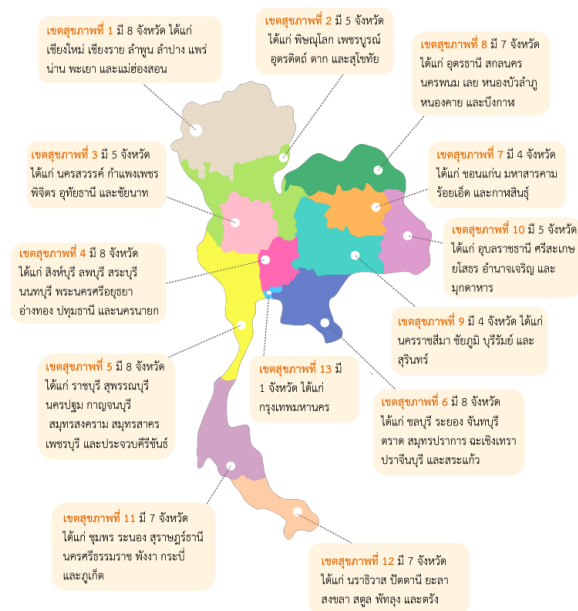
โรคทั้งหมด สูงถึง 702.44 ต่อประชากรหนึ่งแสนคน และโรคนี้มีอัตราการตายตั้งแต่ปี พ.ศ. 2554 ถึง 2559 ที่มีแนวโน้มสูงขึ้น คือ 23.8 23.7 31.1 37.4 42.1 และ 43.8 ต่อประชากรหนึ่งแสนคน ตามลำดับ (Charupronprasit et. al., 2016; Charupronprasit et. al., 2017) ทั้งนี้โรคปอดอักเสบเกิดได้จากการติดเชื้อจากไวรัส แบคทีเรีย และเชื้อรา เชื้อแบคทีเรียที่พบบ่อยได้แก่ เชื้อ Streptococcus Pneumoniae เชื้อ Haemophilus Influenzae Type b เชื้อ Chlamydia Pneumoniae เชื้อ Legionella SPP. และเชื้อ Mycoplasma Pneumoniae เป็นต้น ส่วนเชื้อไวรัส ได้แก่ เชื้อ Respiratory Syncytial Virus (RSV) เชื้อ Influenza หรือเชื้อไข้หวัดใหญ่ เป็นต้น และเชื้อราจะมาจากมูลนกหรือซากพืชซากสัตว์ เมื่อผู้ป่วยติดเชื้อระบบทางเดินหายใจ ระยะพักตัวขึ้นอยู่กับเชื้อก่อโรค อาจใช้เวลา 1 ถึง 3 วัน เมื่อพบผู้ป่วยมีไข้ หอบ อาจมีเจ็บหน้าอก ตรวจร่างกายพบมีเสียง Crepitation ในปอดหรือเสียง Bronchial Breath Sound ในเด็กอายุ 2 เดือนถึง 5 ปี

จากรายงาน Bureau of Epidemiology (2015) เมื่อมีรายงานผู้ป่วยเสียชีวิตเฉียบพลันควรตรวจสอบโรคทันทีเพื่อหาเชื้อก่อโรคและสาเหตุการเสียชีวิต และต้องสอบสวนการระบาดเมื่อพบการป่วยเป็นจำนวนมากเป็นกลุ่มก้อนในพื้นที่เดียวกันเพื่อหาเชื้อก่อโรค ผู้ป่วยสามารถแพร่เชื้อได้จนกว่าเสมหะจากปากและจมูกจะมีเชื้อไม่รุนแรงและปริมาณไม่มากพอ การรักษาผู้ป่วยจะใช้ยาปฏิชีวนะเมื่อวินิจฉัยเป็นการติดเชื้อจากแบคทีเรีย ทั้งนี้หากเป็นการติดเชื้อจากไวรัส แพทย์จะรักษาแบบประคับประคองอาการร่วมกับการบำบัดรักษาระบบทางเดินหายใจให้เหมาะสม

ในปัจจุบันมีการระบาดของเชื้อไวรัสโคโรนา 2019 หรือ โควิด-19 (Covid-19) ซึ่งได้เริ่มแพร่ระบาดเมื่อปลายปี พ.ศ. 2562 และลุกลามไปทั่วโลก สร้างความหวาดกลัวและส่งผลกระทบต่อสุขภาพร่างกายของประชากร ซึ่งโรคระบาดโควิด-19 เป็นโรคที่ส่งผลกระทบต่อระบบทางเดินหายใจและปอดโดยตรง ซึ่งอาจจะทำให้มีผู้ป่วยโรคปอดอักเสบเพิ่มสูงขึ้นตั้งแต่มีการระบาดของโรค ผู้วิจัยจึงมีความสนใจที่จะนำข้อมูลผู้ป่วยโรคปอดอักเสบเพื่อมาพยากรณ์โดยใช้หลักการทางสถิติ

การพยากรณ์โดยใช้หลักการทางสถิติถูกพัฒนามาอย่างต่อเนื่อง เนื่องจากการพยากรณ์มีประโยชน์อย่างมากในการ

วางแผนและการตัดสินใจในการดำเนินงานทั้งระยะสั้นและระยะยาว การพยากรณ์ยังเป็นเครื่องมืออย่างหนึ่งที่จะช่วยให้ได้ข้อมูลในอนาคตเพื่อประกอบการวางแผน จึงเป็นที่ยอมรับกันว่าการพยากรณ์มีบทบาทสำคัญทั้งในงานของภาครัฐและภาคเอกชน ในปัจจุบันตัวแบบพยากรณ์ที่ใช้เทคนิคการวิเคราะห์อนุกรมเวลา หลายวิธีได้ถูกนำมาใช้ในการพยากรณ์ผู้ป่วยทางระบบโรคใน ปี พ.ศ. 2559 Keerativibool (2016) ได้ทำการพยากรณ์รายไตรมาสของผู้ป่วยโรคปอดอักเสบในประเทศไทย โดยใช้วิธีการทางสถิติ 3 วิธี คือ วิธีปรับให้เรียบ วิธีบอกซ์-เจนกินส์ และวิธีพยากรณ์รวม พบว่าผู้ป่วยโรคปอดอักเสบในประเทศไทยมีลักษณะแนวโน้มและมีฤดูกาลรูปแบบบวก ผลสรุปได้วิธีพยากรณ์รวมเป็นวิธีที่มีความเหมาะสมกับอนุกรมเวลาชุดดังกล่าวมากที่สุด ในปี พ.ศ. 2561 Nakunthod & Khamkhod (2018) ได้ศึกษาตัวแบบพยากรณ์อนุกรมเวลารายเดือนของอัตราผู้ป่วยโรคปอดอักเสบในจังหวัดลำปาง ด้วยวิธีการทางสถิติ 4 วิธี คือ วิธีแยกส่วนประกอบ วิธีทำให้เรียบ วิธีบอกซ์-เจนกินส์ และวิธีพยากรณ์รวม และผลสรุปได้วิธีพยากรณ์รวมมีความเหมาะสมกับอนุกรมเวลาชุดดังกล่าวมากที่สุดเช่นกัน



รูปที่ 1 แผนที่เขตสุขภาพ 13 เขตของประเทศไทย

ที่มา: Ekachampaka et. al. (2018)

ดังนั้นในงานวิจัยนี้จึงมีความสนใจที่จะพยากรณ์ผู้ป่วยโรคปอดอักเสบทั่วทั้งประเทศไทย โดยจำแนกตามเขตสุขภาพ ซึ่งในประเทศไทยได้จำแนกเขตสุขภาพเป็นจำนวน 13 เขต ดังรูปที่ 1 โดยงานวิจัยนี้ใช้วิธีการพยากรณ์

ทางสถิติ 4 วิธี ได้แก่ วิธีแยกส่วนประกอบ วิธีปรับให้เรียบ วิธีบ็อกซ์-เจนกินส์ และวิธีพยากรณ์รวม โดยแต่ละวิธีทางสถิติจะมีตัวแบบย่อยแตกต่างกันไปตามแต่ละลักษณะข้อมูลอนุกรมเวลา เพื่อให้ได้วิธีการพยากรณ์ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขต ของประเทศไทยที่เหมาะสมเพื่อผลทำนายที่แม่นยำสามารถนำไปเป็นข้อมูลเบื้องต้นประกอบการวางแผนป้องกัน และควบคุมโรคที่ต้องเฝ้าระวังทางระบาดวิทยาต่อไปทั่วทุกเขตสุขภาพได้

2. วิธีการดำเนินงานวิจัย

การศึกษาวิจัยครั้งนี้มีขั้นตอนวิธีการดำเนินการวิจัยดังนี้

2.1 การจัดเตรียมข้อมูล

การวิจัยนี้มุ่งเน้นเพื่อศึกษาวิธีการพยากรณ์ที่เหมาะสมของผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขตของประเทศไทย ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิที่รวบรวมจาก Health Data Center กระทรวงสาธารณสุข (Health Data Center, Ministry of Public Health, 2023) โดยลักษณะข้อมูลจำแนกเป็นรายเดือน เขตสุขภาพ 1 ถึง 12 จำนวนข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึง เดือนธันวาคม พ.ศ. 2565 ทั้งหมด 96 ค่า เขตสุขภาพ 13 ข้อมูลตั้งแต่เดือนพฤษภาคม พ.ศ. 2561 ถึง เดือนธันวาคม พ.ศ. 2565 ทั้งหมด 56 ค่า โดยใช้จำนวนข้อมูลที่แตกต่างกันตามความสมบูรณ์ของข้อมูล จัดแบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 เป็นชุดข้อมูลฝึกฝน (Training Data Set) ได้แก่ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 1 ถึง 12 ตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2564 จำนวน 84 ค่า ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 ตั้งแต่เดือนพฤษภาคม พ.ศ. 2561 ถึงเดือนธันวาคม พ.ศ. 2564 จำนวน 44 ค่า เพื่อใช้สำหรับการสร้างตัวแบบพยากรณ์ที่เหมาะสมของแต่ละวิธีการพยากรณ์ ส่วนข้อมูลชุดที่ 2 เป็นชุดข้อมูลทดสอบ (Test Data Set) ได้แก่ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 1 ถึง 13 ตั้งแต่เดือนมกราคม พ.ศ. 2565 ถึงเดือนธันวาคม พ.ศ. 2565 จำนวน 12 ค่า โดยใช้เกณฑ์ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบสมมาตร (Symmetric Mean Absolute Percentage Error: sMAPE) เพื่อใช้ตรวจสอบความแม่นยำ (Accuracy) ของตัวแบบพยากรณ์ที่สร้างขึ้น

2.2 องค์ประกอบของอนุกรมเวลา

ในการวิเคราะห์อนุกรมเวลาจะแยกองค์ประกอบต่างๆ ที่ประกอบกันขึ้นเป็นอนุกรมเวลา การทราบองค์ประกอบของอนุกรมเวลาเป็นสิ่งสำคัญในการตัดสินใจเลือกวิธีที่เหมาะสมในการสร้างตัวแบบการพยากรณ์ โดยงานวิจัยนี้ องค์ประกอบของอนุกรมเวลาจะใช้เป็นตัวตัดสินใจในการเลือกตัวแบบการพยากรณ์ในวิธีแยกส่วนประกอบและวิธีปรับให้เรียบ โดยพิจารณาองค์ประกอบของการแปรผันของอนุกรมเวลา 3 ส่วน ดังนี้

2.2.1 แนวโน้ม (Trend) ใช้สถิติทดสอบรันส์ (Runs Test) ในการตรวจสอบ โดยใช้ข้อมูลอนุกรมเวลา (Y) หาค่าเฉลี่ยเคลื่อนที่กลางแบบ 12 เดือนเพื่อกำจัดอิทธิพลของฤดูกาล แล้วใช้ข้อมูลที่กำจัดอิทธิพลของฤดูกาลแล้ว (Y^T) ทดสอบ

2.2.2 การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal Variation) ใช้สถิติทดสอบ Kruskal-Wallis ในการตรวจสอบ โดยใช้ข้อมูลอนุกรมเวลาลบข้อมูลที่กำจัดอิทธิพลของฤดูกาลจะเป็นการกำจัดอิทธิพลของแนวโน้ม $Y^S = Y - Y^T$ แล้วใช้ข้อมูล Y^S ในการทดสอบ

2.2.3 ความแปรปรวนของข้อมูลอนุกรมเวลาใช้สถิติทดสอบของเลวิน (Levene's Test) ถ้าความแปรปรวนของข้อมูลอนุกรมเวลาเท่ากัน (Homogeneity of variance หรือ Homoscedasticity) หรือความแปรปรวนคงที่ จะเลือกใช้สมการรูปแบบบวกและถ้าความแปรปรวนของข้อมูลอนุกรมเวลาไม่เท่ากันจะใช้สมการรูปแบบคูณในการพยากรณ์ข้อมูลอนุกรมเวลาที่มีส่วนประกอบแนวโน้มและฤดูกาล

2.3 การสร้างตัวแบบพยากรณ์

ในการทำวิจัยครั้งนี้ได้กำหนดสัญลักษณ์ไว้ดังนี้ Y_t คือ ข้อมูลอนุกรมเวลา ณ เวลา t
 $\hat{Y}_t$ คือ ค่าพยากรณ์ ณ เวลา t
 ε_t คือ ความคลาดเคลื่อนที่มีการแจกแจงปกติและเป็นอิสระกัน ที่มีเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ทุกช่วงเวลา
 t คือ ช่วงเวลา มีค่าตั้งแต่ 1 ถึง n_1 หรือ n_2 หรือ n
 แล้วแต่กรณี คือ

- กรณีอยู่ในขั้นตอนสร้างตัวแบบ n_1 แทนจำนวนข้อมูลในอนุกรมเวลาชุดข้อมูลฝึกฝน

- กรณีอยู่ในขั้นตอนทดสอบตัวแบบ n_2 แทนจำนวนข้อมูลในอนุกรมเวลาชุดข้อมูลทดสอบและ
- กรณีอยู่ในขั้นตอนการพยากรณ์ล่วงหน้า n แทนจำนวนข้อมูลทั้งหมด

s คือ คาบของฤดูกาลมีค่าเท่ากับ 12 ($s = 12$)

2.3.1 วิธีแยกส่วนประกอบ (Decomposition Method)

วิธีแยกส่วนประกอบเสนอโดย Persons (1919) กล่าวถึงแนวคิดเรื่องการแบ่งอนุกรมเวลาเป็น แนวโน้ม (Trend: T_t) ฤดูกาล (Seasonal: S_t) และส่วนประกอบไม่ปรกติ (Irregular: I_t) ออกจากกัน ส่วนประกอบแนวโน้มสะท้อนถึงทิศทางของข้อมูลระยะยาว ฤดูกาลที่แสดงถึงรูปแบบที่เกิดขึ้นซ้ำๆ กันในช่วงคาบเวลาใดเวลาหนึ่งอย่างเฉพาะเจาะจง และส่วนประกอบไม่ปรกติแสดงถึงความผันผวนของข้อมูลแบบสุ่ม วิธีการนี้มีการพัฒนาอย่างต่อเนื่อง และในปัจจุบันนิยมแยกอนุกรมเวลาออกเป็น ส่วนประกอบต่างๆ 4 ส่วน คือ แนวโน้ม ฤดูกาล วงจร (Cycle: C_t) และส่วนประกอบไม่ปรกติ แต่โดยทั่วไปการพยากรณ์ในระยะสั้น ส่วนประกอบที่จะนำมาพิจารณาในการสร้างตัวแบบพยากรณ์จะมี 2 ส่วนประกอบ คือ แนวโน้มและ/หรือฤดูกาล

ในกรณีที่ส่วนประกอบของข้อมูลอนุกรมเวลาที่จะพิจารณามีเพียงแนวโน้มใช้ ตัวแบบแนวโน้มเส้นตรง (Linear Trend Model) ในการสร้างตัวแบบพยากรณ์ ตัวแบบและตัวแบบพยากรณ์ดังสมการที่ (1) และ (2) ตามลำดับ

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (1)$$

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t \quad (2)$$

เมื่อพารามิเตอร์ β_0, β_1 แทนระดับของข้อมูล และความชัน ตามลำดับ ส่วน $\hat{\beta}_0$ และ $\hat{\beta}_1$ เป็นตัวประมาณของ β_0 และ β_1 ตามลำดับ

ในกรณีที่ส่วนประกอบของข้อมูลอนุกรมเวลาที่จะพิจารณามีแนวโน้มและฤดูกาล จะแบ่งการพิจารณาเป็น 2 รูปแบบ ได้แก่ รูปแบบบวก (Additive Decomposition) เหมาะกับอนุกรมเวลาที่มีความแปรผันของฤดูกาลคงที่ และ รูปแบบคูณ (Multiplicative Decomposition) เหมาะกับอนุกรมเวลาที่มีความแปรผันของฤดูกาลไม่คงที่

รูปแบบบวก ตัวแบบและตัวแบบพยากรณ์ดังสมการที่ (3) และ (4) ตามลำดับ

$$Y_t = \beta_0 + \beta_1 t + S_t + \varepsilon_t \quad (3)$$

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{S}_t \quad (4)$$

เมื่อพารามิเตอร์ S_t แทน ดัชนีวัดความแปรผันเนื่องจากฤดูกาล ณ เวลา t ส่วน $\hat{S}_t$ เป็นตัวประมาณของ S_t (Minsan, 2022)

รูปแบบคูณ ตัวแบบและตัวแบบพยากรณ์ดังสมการที่ (5) และ (6) ตามลำดับ

$$Y_t = (\beta_0 + \beta_1 t) \times S_t + \varepsilon_t \quad (5)$$

$$\hat{Y}_t = (\hat{\beta}_0 + \hat{\beta}_1 t) \times \hat{S}_t \quad (6)$$

2.3.2 วิธีปรับเรียบ (Smoothing Method)

วิธีปรับเรียบนั้นเป็นวิธีการสร้างสมการพยากรณ์จากค่าสังเกตในอนุกรมเวลาบางส่วนหรือทั้งหมดโดยให้น้ำหนักกับค่าสังเกตในอนุกรมเวลาแตกต่างกันแล้วแต่ส่วนประกอบของอนุกรมเวลา

ในกรณีที่ส่วนประกอบของข้อมูลอนุกรมเวลาที่จะพิจารณามีเพียงแนวโน้มใช้ ตัวแบบแนวโน้มเส้นตรงวิธีของ Brown หรือเรียกว่า วิธีทำให้เรียบเอ็กโปเนนเชียลสองครั้ง ในการสร้างตัวแบบพยากรณ์ ตัวแบบพยากรณ์ (7) ดังนี้

$$\hat{Y}_{t+p} = \hat{T}_t + p\hat{B}_t \quad (7)$$

$$\hat{T}_t = 2A_t - A'_t$$

$$\hat{B}_t = \frac{\alpha}{1-\alpha}(A_t - A'_t)$$

เมื่อ A_t คือ ค่าปรับให้เรียบเอ็กโปเนนเชียลครั้งที่ 1 ที่

$$\text{เวลา } t = \alpha Y_t + (1-\alpha)A_{t-1}$$

A'_t คือ ค่าปรับให้เรียบเอ็กโปเนนเชียลครั้งที่ 2 ที่

$$\text{เวลา } t = \alpha A_t + (1-\alpha)A'_{t-1}$$

โดยค่าเริ่มต้นกำหนดได้ดังนี้ $A_1 = Y_1$ และ $A'_1 = Y_1$

α คือ ค่าถ่วงน้ำหนักอยู่ระหว่าง 0 ถึง 1

p คือ จำนวนช่วงเวลาที่ต้องการพยากรณ์ไปล่วงหน้า

ในกรณีที่ส่วนประกอบของข้อมูลอนุกรมเวลาที่จะพิจารณามีแนวโน้มและฤดูกาล จะสร้างตัวแบบพยากรณ์ด้วยวิธีปรับเรียบเส้นโค้งเลขชี้กำลังของโฮลต์-วินเตอร์ รูปแบบบวกหรือแบบคูณ วิธีของโฮลต์-วินเตอร์ ใช้ค่าทำให้เรียบ 3 ค่าได้แก่ α, γ และ δ ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1 สำหรับค่าแนวโน้ม ค่าความลาดชัน และค่าวัดอิทธิพลของฤดูกาลหรือดัชนีฤดูกาลตามลำดับ (Taesombat, 2006)

รูปแบบบวก ตัวแบบพยากรณ์ (8) ดังนี้

$$\hat{Y}_{t+p} = \hat{T}_{t+p} + \hat{S}_{t+p-s} \quad \text{สำหรับ } p=1,2,\dots \quad (8)$$

$$\text{เมื่อ } \hat{T}_{t+p} = \hat{T}_t + p\hat{\beta}_t$$

$\hat{Y}_{t+p}$ คือ ค่าพยากรณ์ ณ เวลา $t+p$

$\hat{T}_t$ คือ ตัวประมาณระดับ $\hat{\beta}_t$ คือ ตัวประมาณแนวโน้ม และ $\hat{S}_t$ คือ ตัวประมาณฤดูกาล ของอนุกรมเวลา ณ เวลา t โดยสามารถคำนวณได้ดังสมการที่ (9) (10) และ (11)

$$\hat{T}_t = \alpha(Y_t - \hat{S}_{t-s}) + (1-\alpha)(\hat{T}_{t-1} + \hat{\beta}_{t-1}) \quad (9)$$

$$\hat{\beta}_t = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1} \quad (10)$$

$$\hat{S}_t = \delta(Y_t - \hat{T}_t) + (1-\delta)\hat{S}_{t-s} \quad (11)$$

รูปแบบคูณ ตัวแบบพยากรณ์ (12) ดังนี้

$$\hat{Y}_{t+p} = \hat{T}_{t+p} \times \hat{S}_{t+p-s} \quad \text{สำหรับ } p=1,2,\dots \quad (12)$$

$$\text{เมื่อ } \hat{T}_{t+p} = \hat{T}_t + p\hat{\beta}_t$$

$\hat{Y}_{t+p}$ คือ ค่าพยากรณ์ ณ เวลา $t+p$

$\hat{T}_t$ คือ ตัวประมาณระดับ $\hat{\beta}_t$ คือ ตัวประมาณแนวโน้ม และ $\hat{S}_t$ คือ ตัวประมาณฤดูกาล ของอนุกรมเวลา ณ เวลา t โดยสามารถคำนวณได้ดังสมการที่ (13) (14) และ (15)

$$\hat{T}_t = \alpha\left(\frac{Y_t}{\hat{S}_{t-s}}\right) + (1-\alpha)(\hat{T}_{t-1} + \hat{\beta}_{t-1}) \quad (13)$$

$$\hat{\beta}_t = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1} \quad (14)$$

$$\hat{S}_t = \delta\left(\frac{Y_t}{\hat{T}_t}\right) + (1-\delta)\hat{S}_{t-s} \quad (15)$$

โดยค่าเริ่มต้นกำหนดได้ดังนี้ $\hat{T}_{initial} = \frac{\sum_{i=1}^s Y_i}{s}$, $\hat{\beta}_{initial} = 0$,

และกรณีรูปแบบบวก $\hat{S}_{initial,j} = Y_j - \hat{T}_{initial}$; $j=1,2,\dots,s$

กรณีรูปแบบคูณ $\hat{S}_{initial,j} = Y_j / \hat{T}_{initial}$; $j=1,2,\dots,s$

2.3.3 วิธีบอกซ์-เจนกินส์ (Box et al., 1994)

เป็นวิธีที่ให้ค่าพยากรณ์ที่มีความถูกต้องสูงกว่าค่าพยากรณ์จากวิธีการอื่นในการพยากรณ์ระยะสั้น การกำหนดตัวแบบพยากรณ์ทำได้โดยการตรวจสอบฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) ของอนุกรมเวลาแบบคงที่ (Stationary Time Series) หรืออนุกรมเวลาที่มีค่าเฉลี่ยของความแปรปรวนคงที่ (Taesombat, 2006) ตัวแบบทั่วไปวิธีบอกซ์-เจนกินส์ คือ Seasonal Autoregressive Integrated

Moving Average หรือ SARIMA(p,d,q)(P,D,Q)_s ดังที่แสดงใน (16)

$$\phi_p(\mathbf{B})\Phi_P(\mathbf{B}^s)(1-\mathbf{B})^d(1-\mathbf{B}^s)^D Y_t = \delta + \theta_q(\mathbf{B})\Theta_Q(\mathbf{B}^s)\varepsilon_t \quad (16)$$

เมื่อ $\phi_p(\mathbf{B}) = 1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 - \phi_3\mathbf{B}^3 \dots - \phi_p\mathbf{B}^p$ แทนการ

ถดถอยในตัวแบบไม่มีฤดูกาลอันดับ p (Non-seasonal Autoregressive Operator of Order p : AR(p))

$\Phi_P(\mathbf{B}^s) = 1 - \Phi_1\mathbf{B}^s - \Phi_2\mathbf{B}^{2s} - \Phi_3\mathbf{B}^{3s} \dots - \Phi_P\mathbf{B}^{Ps}$ แทน

การถดถอยในตัวแบบมีฤดูกาลอันดับ P (Seasonal Autoregressive Operator of Order P : SAR(P))

$(1-\mathbf{B})^d$ คือ อันดับของผลต่างแบบไม่มีฤดูกาล (Non-seasonal Difference)

$(1-\mathbf{B}^s)^D$ คือ อันดับของผลต่างแบบมีฤดูกาล (Seasonal Difference)

$\delta = \phi_p(\mathbf{B})\Phi_P(\mathbf{B}^s)\mu$ คือ ค่าคงตัว โดยที่ μ คือ ค่าเฉลี่ยของอนุกรมเวลาที่คงที่

$\theta_q(\mathbf{B}) = 1 - \theta_1\mathbf{B} - \theta_2\mathbf{B}^2 - \theta_3\mathbf{B}^3 \dots - \theta_q\mathbf{B}^q$ แทนค่าเฉลี่ย

เคลื่อนที่แบบไม่มีฤดูกาลอันดับ q (Non-seasonal Moving Average Operator of Order q : MA(q))

$\Theta_Q(\mathbf{B}^s) = 1 - \Theta_1\mathbf{B}^s - \Theta_2\mathbf{B}^{2s} - \Theta_3\mathbf{B}^{3s} \dots - \Theta_Q\mathbf{B}^{Qs}$ แทน

ค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ Q (Seasonal Moving Average Operator of Order Q : SMA(Q))

$\mathbf{B}$ คือ ตัวดำเนินการถอยหลัง (Backward Operator)

โดยที่ $\mathbf{B}^s Y_t = Y_{t-s}$

d และ D คือ ลำดับที่ของการหาผลต่างและผลต่าง

ฤดูกาลของอนุกรมเวลา ตามลำดับ

t คือ ช่วงเวลา มีค่าตั้งแต่ 1 ถึง n_1 หรือ n_2 หรือ n แล้วแต่กรณี

การสร้างตัวแบบพยากรณ์โดยวิธี บอกซ์-เจนกินส์ มีขั้นตอนดังนี้

1) การตรวจสอบข้อมูลเพื่อพิจารณาว่าอนุกรมเวลาคงที่หรือไม่ โดยพิจารณาจากกราฟอนุกรมเวลาและพิจารณาจากกราฟ ACF และ PACF และการตรวจสอบ Unit Root ด้วยการทดสอบ Dickey-Fuller (Dickey-Fuller Test) ในงานวิจัยนี้กำหนดลำดับมากที่สุดของ Lag = 12 เนื่องมาจากการเลือกลำดับ Lag ที่เหมาะสมคือค่าต่ำที่สุดของเกณฑ์สารสนเทศของอะไคเกะ (Akaike's Information Criterion: AIC) หากพบว่าอนุกรมเวลาไม่คงที่ (Non-

stationary) ต้องแปลงอนุกรมเวลาให้คงที่ก่อนที่จะดำเนินการในขั้นตอนต่อไป เช่น การแปลงข้อมูลด้วยการหาผลต่างและ/หรือผลต่างฤดูกาล (Regular Difference or Seasonal Difference) เมื่ออนุกรมเวลามีแนวโน้มและ/หรือฤดูกาล การแปลงข้อมูลแบบ Box-Cox เมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่เป็นต้น

2) การกำหนดตัวแบบพยากรณ์ที่เป็นไปได้จากกราฟ ACF และ PACF ของอนุกรมเวลาใหม่ที่อนุกรมเวลาแบบคงที่ นั่นคือ กำหนดค่า p, q, P และ Q พร้อมทั้งประมาณค่าพารามิเตอร์ของตัวแบบด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method)

3) การตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ที่กำหนด โดยการพิจารณาคัดเลือกตัวแบบพยากรณ์ที่มีค่าเกณฑ์ AIC ที่ต่ำที่สุด ทดสอบตัวประมาณค่าพารามิเตอร์ของตัวแบบโดยสถิติการทดสอบที (t-test) และตรวจสอบข้อสมมติเกี่ยวกับความคลาดเคลื่อนจากการพยากรณ์ (e_t) ที่ต้องมีคุณลักษณะความเป็นอิสระกันและมีการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ทุกช่วงเวลา นั่นคือ

- ความเป็นอิสระกัน โดยสถิติการทดสอบ Ljung-Box เมื่อกำหนดจำนวน Lag ที่พิจารณาเท่ากับ 36 (LB Lag36)
- การแจกแจงปกติ โดยสถิติการทดสอบคอลโมโกรอฟ-สมิรโนฟ (Kolmogorov-Smirnov Test: KS)
- ค่าเฉลี่ยเท่ากับศูนย์ โดยสถิติการทดสอบที (t-test)
- ความแปรปรวนคงที่ทุกช่วงเวลา โดยสถิติการทดสอบของเลวี (Levene's Test: Levene)

4) พยากรณ์อนุกรมเวลาโดยใช้ตัวแบบพยากรณ์ที่เหมาะสมที่สุดจากขั้นตอนที่ 3

2.3.4 การสร้างตัวแบบพยากรณ์โดยวิธีพยากรณ์รวม (Combined Forecasting Method)

การพยากรณ์รวมเป็นวิธีการประยุกต์ที่มีการรวมค่าพยากรณ์จากวิธีการพยากรณ์เดี่ยวตั้งแต่ 2 วิธีขึ้นไปเพื่อให้ได้ค่าพยากรณ์ใหม่ที่มีความคลาดเคลื่อนน้อยที่สุด การพยากรณ์รวมมีด้วยกันหลายวิธี เช่น วิธีถ่วงน้ำหนักความ

คลาดเคลื่อนกำลังสองเฉลี่ยผกผัน (Inverse of Mean Squares Error) วิธีถ่วงน้ำหนักความคลาดเคลื่อนสมบูรณ์เฉลี่ยต่ำสุด (Min Mean Absolute Error) และ วิธีหาค่าเหมาะที่สุดแบบวาฬ (Whale Optimization Algorithm) (Minsan et al., 2021; Minsan & Minsan, 2023) และวิธีถ่วงน้ำหนักด้วยสัมประสิทธิ์การถดถอย (Regression Analysis: REG) เป็นต้น Keerativibool (2016) และ Nakunthod & Khamkhod (2018) ที่ได้ศึกษาตัวแบบพยากรณ์อนุกรมเวลาผู้ป่วยโรคปอดอักเสบ พบว่าการพยากรณ์รวมด้วยวิธี REG เป็นวิธีการที่เหมาะสมสำหรับทั้ง 2 งานวิจัย โดยมีข้อแตกต่างที่ Keerativibool (2016) จะนำค่าสัมประสิทธิ์การถดถอยไปคำนวณค่าถ่วงน้ำหนักก่อนนำไปใช้ แต่ Nakunthod & Khamkhod (2018) จะนำค่าสัมประสิทธิ์การถดถอยที่คำนวณได้ไปใช้ในการพยากรณ์รวมด้วยลักษณะการถดถอยพหุคูณ (Multiple Regression) และเมื่อศึกษางานวิจัยอื่นเพิ่มเติมเกี่ยวกับการพยากรณ์รวมที่นำไปใช้พยากรณ์ในข้อมูลหลากหลาย

Saeying et. al. (2023) พบว่าตัวแบบพยากรณ์ที่เหมาะสมสำหรับพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำของเขื่อนการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.) 11 แห่ง วิธีพยากรณ์รวม REG เป็นวิธีที่เหมาะสมกับอนุกรมเวลาปริมาณน้ำไหลเข้าอ่างเก็บน้ำของเขื่อนทั้ง 11 แห่ง Tunkaew et. al. (2023) หาตัวแบบพยากรณ์ที่เหมาะสมสำหรับพยากรณ์อัตราแลกเปลี่ยนสกุลเงินต่างประเทศ 10 สกุลที่คนไทยนิยมไปท่องเที่ยวมากที่สุดเทียบกับค่าเงินบาท พบว่าวิธีพยากรณ์รวม REG เป็นวิธีการการพยากรณ์ที่เหมาะสมเป็นส่วนใหญ่ ดังนั้นในงานวิจัยนี้จึงเลือกใช้วิธีพยากรณ์รวม REG ด้วยลักษณะการถดถอยพหุคูณ (Multiple Regression) ตัวแบบพยากรณ์ในสมการที่ (17) ดังนี้

$$\hat{Y}_t = b_0 + b_1 \hat{Y}_{1t} + b_2 \hat{Y}_{2t} + b_3 \hat{Y}_{3t} \quad (17)$$

เมื่อ $\hat{Y}_t$ คือ ค่าพยากรณ์รวม ณ เวลา t และ $\hat{Y}_{1t}, \hat{Y}_{2t}$ และ $\hat{Y}_{3t}$ แทนค่าพยากรณ์เดี่ยว ณ เวลา t จากวิธีแยกส่วนประกอบ วิธีทำให้เรียบ และวิธีบอกซ์-เจนกินส์ตามลำดับ

b_0 คือ ตัวประมาณค่าคงตัวของตัวแบบการวิเคราะห์ถดถอย

b_1, b_2 และ b_3 แทนตัวประมาณค่าถ่วงน้ำหนักของแต่ละวิธีพยากรณ์ด้วยวิธีกำลังสองน้อยที่สุด (Least Squares Method) (Montgomery et al., 2006)

2.3.5 การเปรียบเทียบความถูกต้องของค่าพยากรณ์ด้วยร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบสมมาตร (Symmetry Mean Absolute Percent Error: sMAPE)

ความถูกต้องของการพยากรณ์เป็นสิ่งที่ผู้ใช้ค่าพยากรณ์ต้องการ ความถูกต้องมีมากหรือน้อยขึ้นอยู่กับค่าความคลาดเคลื่อนของการพยากรณ์ (Forecast Error: e_t) ซึ่งเป็นผลต่างของค่าจริงและค่าพยากรณ์ ($e_t = Y_t - \hat{Y}_t$) ความคลาดเคลื่อนจะมากถ้าค่าจริงห่างจากค่าพยากรณ์มาก และจะน้อยถ้าค่าพยากรณ์ใกล้เคียงกับค่าจริง ในที่นี้เลือกการพิจารณาค่าความแม่นยำของตัวแบบพยากรณ์ใช้ sMAPE (Armstrong, 1985) (Flores, 1986) ซึ่งเป็นค่าที่ไม่มีหน่วย แสดงผลลัพธ์เป็นร้อยละ เป็นวิธีที่เหมาะสมเพื่อใช้เปรียบเทียบอนุกรมเวลาหลายชุดเมื่อใช้วิธีพยากรณ์เดียวกัน และ/หรือเปรียบเทียบวิธีการพยากรณ์หลายวิธีเมื่อใช้อนุกรมเวลาชุดเดียวกัน และมีความเหมาะสมสำหรับการพยากรณ์ที่มีข้อมูลจริง (Actual Data) มีค่าบางค่าที่ค่าเป็นศูนย์หรือใกล้เคียงศูนย์ สมการแสดงใน (18)

$$sMAPE = \frac{100}{n_2} \sum_{t=1}^{n_2} \frac{|e_t|}{(Y_t + \hat{Y}_t)/2}, \quad (18)$$

(Hyndman & Koehler, 2006)

เมื่อ n_2 แทนจำนวนข้อมูลในอนุกรมเวลาชุดข้อมูลทดสอบ ($n_2 = 12$)

3. ผลการวิจัย

จากการพิจารณาลักษณะการการเคลื่อนไหวของอนุกรมเวลาผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขตของชุดข้อมูลฝึกฝนตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือน ธันวาคม พ.ศ. 2564 จำนวน 84 ค่า สำหรับเขตสุขภาพที่ 1-12 และจำนวน 44 ค่า สำหรับเขตสุขภาพที่ 13 ได้ผลการพิจารณาการเคลื่อนไหวของอนุกรมเวลาและการวิเคราะห์ด้วยวิธีการพยากรณ์ 4 วิธี มีรายละเอียดดังนี้

3.1. การพิจารณาการเคลื่อนไหวของอนุกรมเวลา

ผลการวิเคราะห์ลักษณะการเคลื่อนไหวของอนุกรมเวลา ทดสอบองค์ประกอบแนวโน้ม ฤดูกาล และความแปรปรวนของข้อมูล โดยใช้สถิติทดสอบรันส์ ทดสอบ Kruskal-Wallis และทดสอบของเลวิน ตามลำดับ แสดง

ในตารางที่ 1 พบว่าข้อมูลผู้ป่วยโรคปอดอักเสบเขตสุขภาพที่ 1 ถึง 12 สถิติทดสอบรันส์และ Kruskal-Wallis ค่าพินัยกว่าระดับนัยสำคัญที่กำหนด แสดงว่าข้อมูลอนุกรมเวลามีอิทธิพลจากแนวโน้มและฤดูกาล ส่วนเขตสุขภาพที่ 13 สถิติทดสอบรันส์ค่าพินัยกว่าระดับนัยสำคัญที่กำหนด แต่สถิติทดสอบ Kruskal-Wallis ค่าพินัยมากกว่าระดับนัยสำคัญที่กำหนด ดังนั้นเขตสุขภาพที่ 13 จึงมีเฉพาะแนวโน้ม

เมื่อพิจารณาสถิติทดสอบของเลวินค่าพินัยมากกว่าระดับนัยสำคัญที่กำหนดในเขตสุขภาพที่ 1, 2, 8, 9, 10, 11, 12 และ 13 แสดงว่า ข้อมูลผู้ป่วยโรคปอดอักเสบมีความแปรปรวนคงที่ ดังนั้นจึงกำหนดให้เขตสุขภาพที่ 1, 2, 8, 9, 10, 11 และ 12 มีอิทธิพลจากแนวโน้มและฤดูกาลเป็นตัวแบบรูปแบบบวก ข้อมูลผู้ป่วยโรคปอดอักเสบที่มีความแปรปรวนไม่คงที่ประกอบด้วย 5 เขต คือ เขตสุขภาพที่ 3 ถึง 7 กำหนดให้มีอิทธิพลจากแนวโน้มและฤดูกาลเป็นตัวแบบรูปแบบคูณ ดังนั้นผลสรุปเขตสุขภาพที่ 1 ถึง 12 จะไปกำหนดตัวแบบในวิธีแยกส่วนประกอบที่มีอิทธิพลจากแนวโน้มและฤดูกาล และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังของโฮลต์-วินเทอร์

สุดท้ายเขตสุขภาพที่ 13 ที่มีอิทธิพลจากแนวโน้มเพียงอย่างเดียวจะใช้วิธีแยกส่วนประกอบที่มีเฉพาะแนวโน้มเส้นตรงและวิธีทำให้เรียบเอ็กโปเนนเชียลสองครั้งในการสร้างตัวแบบ

3.2. การวิเคราะห์อนุกรมเวลาโดยวิธีแยกส่วนประกอบ

การคำนวณหาความแปรผันฤดูกาล กรณีที่เป็นตัวแบบบวกความแปรผันของฤดูกาลจะแสดงด้วยค่าดัชนีวัดความแปรผันเนื่องจากฤดูกาลที่มีเกณฑ์เทียบจำนวนผู้ป่วยโรคปอดอักเสบปกติจะอยู่ที่ค่าเท่ากับ 0 กรณีที่เป็นตัวแบบคูณความแปรผันของฤดูกาลจะแสดงด้วยค่าดัชนีวัดความแปรผันเนื่องจากฤดูกาลที่มีเกณฑ์เทียบจำนวนผู้ป่วยโรคปอดอักเสบปกติจะอยู่ที่ค่าเท่ากับ 1 แสดงในตารางที่ 2 และตัวแบบเส้นตรงแสดงในตารางที่ 3

จากตารางที่ 2 พบว่าอิทธิพลของฤดูกาลของแต่ละเขตสุขภาพที่ 1 ถึง 12 มีลักษณะคล้ายกันที่ค่าต่ำสุดและค่าสูงสุด นั่นคือจำนวนผู้ป่วยโรคปอดอักเสบจะมีจำนวนน้อยที่สุดทุกเขตสุขภาพในช่วงเดือนธันวาคมและมีจำนวนมากที่สุดในช่วงฤดูฝนคือประมาณเดือนสิงหาคม

ถึง ตุลาคม และมีจำนวนผู้ป่วยค่อนข้างสูงในช่วงฤดูหนาว คือประมาณเดือน มกราคม ถึง กุมภาพันธ์

3.3 การสร้างตัวแบบพยากรณ์โดยวิธีปรับให้เรียบ

การสร้างตัวแบบพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังของโฮลต์-วินเทอร์รูปแบบบวกหรือคูณสามารถประมาณค่าพารามิเตอร์ของอนุกรมเวลาผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 1 ถึง 12 เขต ได้ค่า α , γ และ δ และวิธีทำให้เรียบอีกโปเนนเชียลสองครั้งในเขตสุขภาพ 13 ได้ค่า α ในตารางที่ 4

3.4 การสร้างตัวแบบพยากรณ์โดยวิธีบอซ-เจนกินส์

ตัวแบบพยากรณ์โดยวิธีบอซ-เจนกินส์ ของอนุกรมเวลาผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขต ประกอบไปด้วยตัวแบบ $SARIMA(p,d,q)(P,D,Q)_{12}$ แตกต่างกันแสดงในตารางที่ 5 และ 6

จากตารางที่ 6 พบว่า การทดสอบตัวประมาณค่าพารามิเตอร์ของตัวแบบโดยสถิติการทดสอบที ผ่านทุกตัวแบบ และจากตารางที่ 5 พบว่า ข้อมูลอนุกรมเวลาผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 1 2 3 5 7 8 9 10 11 และ 12 ได้ตัวแบบ $SARIMA(p,d,q)(P,D,Q)_{12}$ ที่เหมาะสมผ่านทุกเงื่อนไข มีเพียงเขตสุขภาพที่ 4 ที่ไม่ผ่านเงื่อนไขความคลาดเคลื่อนมีความแปรปรวนคงที่ เขตสุขภาพที่ 6 ที่ไม่ผ่านเงื่อนไขความคลาดเคลื่อนมีการแจกแจงปกติ และ เขตสุขภาพที่ 13 ที่ไม่ผ่านเงื่อนไขความคลาดเคลื่อนมีความแปรปรวนคงที่และความคลาดเคลื่อนมีการแจกแจงปกติ ดังนั้นตัวแบบเขตสุขภาพที่ 4 6 และ 13 จะไม่นำไปใช้ในการพยากรณ์

3.5 การสร้างตัวแบบการพยากรณ์โดยวิธีพยากรณ์รวม

การใช้ตัวแบบพยากรณ์ 3 วิธีแรกมาสร้างตัวแบบการพยากรณ์รวม โดยใช้วิธีถ่วงด้วยสัมประสิทธิ์การถดถอย (REG) สามารถหาค่าถ่วงน้ำหนักของแต่ละวิธีได้ผลในตารางที่ 7

3.6 การเปรียบเทียบวิธีการพยากรณ์

จากการพยากรณ์ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขต ของประเทศไทย ทั้ง 4 วิธีได้ค่าพยากรณ์รายเดือน ตั้งแต่เดือน มกราคม พ.ศ. 2565 ถึง เดือน ธันวาคม พ.ศ.

2565 แล้วนำมาคำนวณค่า sMAPE แต่ละวิธี เพื่อเปรียบเทียบตัวแบบพยากรณ์ที่เหมาะสมได้ผลในตารางที่ 8

จากตารางที่ 8 จะเลือกวิธีการพยากรณ์ที่ให้ค่า sMAPE ต่ำที่สุด แต่ถ้าหากมีกรณีค่า sMAPE ของการพยากรณ์เดี่ยว และพยากรณ์รวมได้ค่าเท่ากันจะเลือกใช้วิธีการพยากรณ์เดี่ยว สรุปผลได้ว่า ส่วนใหญ่แล้วตัวแบบที่มีความเหมาะสมจะใช้เทคนิคตัวแบบพยากรณ์รวมโดยใช้วิธี REG ใช้ตัวแบบนี้ด้วยกันทั้งหมด 11 เขต คือ เขตสุขภาพที่ 1 ถึง 6 และเขตสุขภาพที่ 8 ถึง 10 และเขตสุขภาพที่ 12 และ 13 และตัวแบบพยากรณ์ที่เหมาะสมของเขตสุขภาพที่ 7 คือ วิธีแยกส่วนประกอบที่มีส่วนประกอบของแนวโน้ม และฤดูกาลรูปแบบคูณและเขตสุขภาพที่ 11 ตัวแบบพยากรณ์ที่เหมาะสมคือ ตัวแบบพยากรณ์บอซ-เจนกินส์

3.7 การพยากรณ์ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขต ของประเทศไทย

ในการพยากรณ์ผู้ป่วยโรคปอดอักเสบในเขตสุขภาพ 13 เขต ของประเทศไทยรายเดือนในเดือน มกราคม พ.ศ. 2566 ถึง เดือนธันวาคม พ.ศ. 2566 เลือกใช้วิธีการพยากรณ์ที่ตัวแบบพยากรณ์ให้ค่า sMAPE น้อยที่สุดในแต่ละเขตสุขภาพ ได้ผลการพยากรณ์แสดงในตารางที่ 9 และรูปที่ 2 และ 3

จากรูปที่ 2 และ 3 พบว่าผู้ป่วยโรคปอดอักเสบในแต่ละเขตสุขภาพ 1 ถึง 12 นั้นมีลักษณะฤดูกาลของโรคคล้ายคลึงกัน คือมีลักษณะฤดูกาลเป็นรูปตัวอักษร M โดยจะมีช่วงผู้ป่วยมากในช่วงปลายฤดูหนาวคาบเกี่ยวไปต้นฤดูร้อน และ ช่วงปลายฤดูฝนคาบเกี่ยวไปต้นฤดูหนาว นั่นคือช่วงเวลาที่เปลี่ยนผ่านของฤดูกาลจะเป็นเหตุให้มีผู้ป่วยโรคปอดอักเสบมาก อย่างไรก็ตามไม่สามารถใช้เหตุผลนี้สรุปในเขตสุขภาพที่ 13 คือจังหวัดกรุงเทพมหานครได้ จังหวัดนี้ไม่พบลักษณะที่เป็นฤดูกาล เป็นจังหวัดที่มีความความผันผวนของจำนวนผู้ป่วยที่คาดการณ์ได้ยาก (ค่า sMAPE สูง) แต่ตัวแบบพยากรณ์ก็พยากรณ์ให้เห็นถึงแนวโน้มที่เพิ่มขึ้นของจำนวนผู้ป่วย จึงจำเป็นต้องเฝ้าระวังในเขตสุขภาพนี้

ตารางที่ 1 สถิติทดสอบและค่าของสถิติทดสอบรันส์ Kruskal-Wallis และของเลวีเน่ แยกตามเขตสุขภาพ 13 เขต

Health Districts	Runs Test		Kruskal-Wallis Test		Levene's Test	
	Test Statistic	P-value	Test Statistic	P-value	Test Statistic	P-value
1	-8.071	0.000*	54.84	0.000*	0.004	0.952
2	-8.071	0.000*	50.26	0.000*	0.036	0.850
3	-7.121	0.000*	46.37	0.000*	4.219	0.043*
4	-7.834	0.000*	37.54	0.000*	6.891	0.010*
5	-7.596	0.000*	47.82	0.000*	4.243	0.043*
6	-7.359	0.000*	43.36	0.000*	5.252	0.024*
7	-7.596	0.000*	47.45	0.000*	9.224	0.003*
8	-7.359	0.000*	37.58	0.000*	2.844	0.096
9	-7.121	0.000*	50.91	0.000*	0.861	0.356
10	-7.596	0.000*	35.96	0.000*	0.269	0.605
11	-8.071	0.000*	54.11	0.000*	0.041	0.839
12	-7.359	0.000*	46.49	0.000*	1.964	0.165
13	-5.391	0.000*	9.56	0.570	3.12	0.085

หมายเหตุ * ปฏิเสธสมมติฐานว่างที่ระดับนัยสำคัญ 0.05

ตารางที่ 2 ค่าประมาณดัชนีวัดความแปรผันเนื่องจากฤดูกาลด้วยวิธีแยกส่วนประกอบแยกตามเขตสุขภาพ 13 เขต

Health Districts	Seasonal Indices											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1 ^A	1,855	1,722	2,043	-105	-1,243	-1,511	-1,737	-30	1,653	2,730	983	-6,360
2 ^A	818	1,474	889	-726	-852	-837	-519	890	1,999	1,404	-201	-4,340
3 ^B	1.24	1.21	1.21	0.90	0.73	0.76	0.93	1.00	1.27	1.31	1.01	0.43
4 ^B	1.17	1.16	1.15	0.78	0.72	0.88	0.93	1.14	1.53	1.34	0.93	0.27
5 ^B	1.21	1.19	1.08	0.79	0.80	0.88	0.96	1.10	1.38	1.32	0.97	0.32
6 ^B	1.22	1.04	1.07	0.79	0.77	0.89	1.09	1.36	1.37	1.30	0.93	0.17
7 ^B	1.14	1.18	1.21	0.86	0.81	0.84	0.85	1.09	1.29	1.43	1.12	0.19
8 ^A	547	525	989	-1,087	-465	-667	-527	811	1,813	1,113	-281	-2,769
9 ^A	2,238	2,927	1,978	-1,256	-2,021	-1,621	-2,065	704	2,257	4,191	-224	-7,109
10 ^A	1,084	1,324	1,808	-1,808	-2,177	-1,714	-284	351	4,604	3,498	-223	-6,461
11 ^A	1,497	741	476	-1,581	-1,560	-1,141	-184	1,039	3,136	1,655	-363	-3,715
12 ^A	1,414	170	-23	-1,706	-1,812	-784	10	1,563	2,492	1,322	264	-2,910
13	-	-	-	-	-	-	-	-	-	-	-	-

หมายเหตุ ค่าน้อย  ค่ามากสำหรับแต่ละเขตสุขภาพ^A ส่วนประกอบแนวโน้มและฤดูกาลรูปแบบบวก^B ส่วนประกอบแนวโน้มและฤดูกาลรูปแบบคูณตารางที่ 3 ค่าประมาณพารามิเตอร์ของตัวแบบเส้นตรง $\hat{\beta}_0$ และ $\hat{\beta}_1$ วิธีแยกส่วนประกอบแยกตามเขตสุขภาพ 13 เขต

Health Districts	Parameters	
	$\hat{\beta}_0$	$\hat{\beta}_1$
1	8,201.0	0.40
2	6,713.0	-45.39
3	3,109.0	-5.99
4	-2,948.0	312.50
5	5,714.0	-18.15
6	4,434.0	88.80
7	14,899.0	-70.40
8	7,403.0	-47.64
9	12,581.0	-58.10
10	12,424.0	-41.00
11	5,616.0	-5.99
12	4,618.0	15.81
13	28.0	9.23

ตารางที่ 4 ค่าพารามิเตอร์ α , γ และ δ วิธีปรับเรียบแยกตามเขตสุขภาพ 13 เขต

Health Districts	Parameters		
	α	γ	δ
1	0.54	0.00	0.46
2	0.94	0.00	1.00
3	0.21	0.00	0.58
4	0.36	0.41	0.97
5	0.15	0.00	0.48
6	0.00	1.00	0.00
7	0.00	0.03	0.78
8	0.85	0.00	1.00
9	0.89	0.00	1.00
10	0.80	0.00	1.00
11	0.91	0.00	0.84
12	0.65	0.00	1.00
13	0.07	-	-

หมายเหตุ ค่าน้อย  ค่ามากสำหรับแต่ละเขตสุขภาพ

0.00 ใช้แสดงแทนเมื่อค่าเข้าใกล้ศูนย์เกินทศนิยมตำแหน่งที่ 2

1.00 ใช้แสดงแทนเมื่อค่าเข้าใกล้หนึ่งเกินทศนิยมตำแหน่งที่ 2

ตารางที่ 5 ตัวแบบบอกร์-เจนกินส์ และการตรวจสอบความเหมาะสมของตัวแบบแยกตามเขตสุขภาพ 13 เขต

Health Districts	Box-Jenkins	t-test P-value	Levene P-value	KS P-value	LB Lag 36 P-value
1	SARIMA(2,1,2)(0,1,0) ₁₂	-0.218 0.828	0.537 0.466	0.094 0.120	20.517 0.942
2	SARIMA(1,1,2)(0,1,0) ₁₂	-0.856 0.395	2.597 0.112	0.103 0.061	38.162 0.246
3	SARIMA(0,1,0)(0,1,1) ₁₂	0.128 0.899	1.573 0.214	0.089 >0.150	30.220 0.698
4	SARIMA(2,1,2)(0,1,0) ₁₂ ^A	-0.302 0.763	8.892 0.004*	0.065 >0.150	24.115 0.840
5	SARIMA(0,1,2)(1,1,0) ₁₂	-1.081 0.283	1.012 0.318	0.088 >0.150	33.987 0.420
6	SARIMA(1,0,0)(0,1,1) ₁₂ ^{A, B}	0.452 0.653	0.439 0.510	0.193 0.010*	15.735 0.995
7	SARIMA(2,1,0)(0,1,1) ₁₂	-0.181 0.857	0.809 0.372	0.073 >0.150	27.373 0.743
8	SARIMA(0,1,0)(2,1,0) ₁₂	0.078 0.938	1.557 0.216	0.078 >0.150	44.357 0.110
9	SARIMA(0,1,0)(0,1,1) ₁₂	-0.128 0.899	0.776 0.381	0.081 >0.150	46.500 0.093
10	SARIMA(1,0,0)(0,1,1) ₁₂	0.052 0.959	0.012 0.913	0.087 >0.150	37.110 0.328
11	SARIMA(1,0,0)(0,1,1) ₁₂	0.250 0.803	0.235 0.629	0.096 0.098	47.241 0.065
12	SARIMA(1,1,1)(0,1,0) ₁₂	-0.237 0.814	2.770 0.101	0.100 0.077	28.723 0.724
13	SARIMA(1,0,1)(2,0,0) ₁₂ ^B	-0.042 0.967	16.760 0.000*	0.324 <0.010*	21.163 0.908

หมายเหตุ * ปฏิเสธสมมติฐานว่างที่ระดับนัยสำคัญ 0.05

^A Box-Cox transformation by natural log

^B the model that include a constant term.

ตารางที่ 6 สถิติทดสอบที่ ค่าคงตัว AR MA SAR และ SMA ของตัวแบบบอซ-เจนกินส์ แยกตามเขตสุขภาพ 13 เขต

Health Districts	SARIMA(p, d, q)(P, D, Q) ₁₂											
	Ln	d=1	D=1	Constant P-value	AR(1) P-value	AR(2) P-value	MA(1) P-value	MA(2) P-value	SAR(1) P-value	SAR(2) P-value	SMA(1) P-value	SMA(2) P-value
1	No	Yes	Yes		12.800 0.000*	-11.200 0.000*	15.900 0.000*	-15.600 0.000*				
2	No	Yes	Yes		2.300 0.024*		2.900 0.005*	2.200 0.029*				
3	No	Yes	Yes								7.400 0.000*	
4	Yes	Yes	Yes		7.200 0.000*	-3.700 0.000*	8.900 0.000*	-10.100 0.000*				
5	No	Yes	Yes				2.100 0.040*	3.700 0.000*	-5.900 0.000*			
6	Yes	No	Yes	2.400 0.018*	7.800 0.000*						6.300 0.000*	
7	No	Yes	Yes		2.100 0.037*	-3.800 0.000*					7.400 0.000*	
8	No	Yes	Yes						-7.000 0.000*	-2.700 0.008*		
9	No	Yes	Yes								5.900 0.000*	
10	No	No	Yes		8.700 0.000*						7.100 0.000*	
11	No	No	Yes		13.500 0.000*						6.900 0.000*	
12	No	Yes	Yes		6.800 0.000*		18.400 0.000*					
13	No	No	No	10.200 0.000*	-2.200 0.031*		3.700 0.001*		-6.000 0.000*	-6.900 0.000*		

หมายเหตุ * ปฏิเสธสมมติฐานว่างที่ระดับนัยสำคัญ 0.05

ตารางที่ 7 การประมาณค่าพารามิเตอร์ b_0 , b_1 , b_2 and b_3 ของวิธี REG แยกตามเขตสุขภาพ 13 เขต

Health Districts	Parameters			
	b_0	b_1	b_2	b_3
1	-903.06	0.29	0.21	0.62
2	110.53	0.35	-0.03	0.68
3	132.51	0.28	0.10	0.58
4	760.53	0.21	0.71	-
5	147.32	0.25	0.04	0.69
6	-412.99	1.69	-0.61	-
7	316.36	0.41	-0.15	0.74
8	-103.28	0.33	0.07	0.63
9	84.50	0.53	-0.44	0.92
10	-292.40	0.22	0.02	0.80
11	199.35	0.05	-0.19	1.13
12	-	0.65	0.54	0.01
13	2.56	0.87	0.12	-

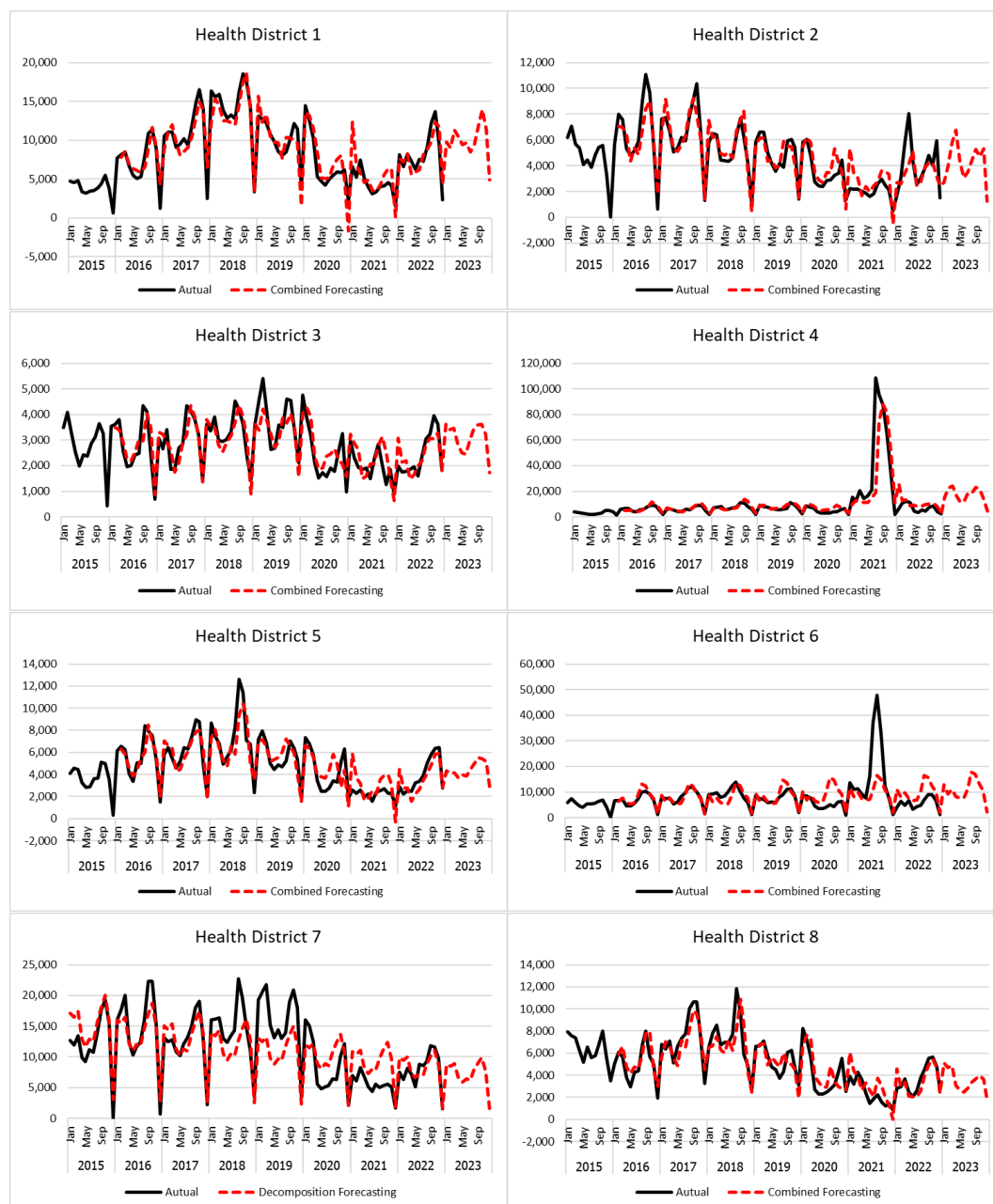
ตารางที่ 8 การเปรียบเทียบค่า sMAPE ของวิธีแยกส่วนประกอบ การปรับเรียบ บอซ-เจนกินส์ และการพยากรณ์รวมแยกตามเขตสุขภาพ 13 เขต

Health Districts	sMAPE			
	Decomposition	Smoothing	Box-Jenkins	Combined
1	17.15	47.40	74.91	16.61*
2	61.87	50.34	115.59	27.62*
3	27.16	35.71	32.93	16.99*
4	114.32	106.28	-	47.20*
5	31.46	51.95	115.42	20.59*
6	76.31	100.21	-	53.43*
7	16.15*	31.13	36.18	18.97
8	39.80	34.88	75.36	15.01*
9	35.87	36.17	32.88	28.37*
10	33.83	33.92	23.00	19.40*
11	24.89	30.01	16.75*	19.92
12	30.84	87.41	55.87	23.51*
13	105.70	104.49	-	97.84*

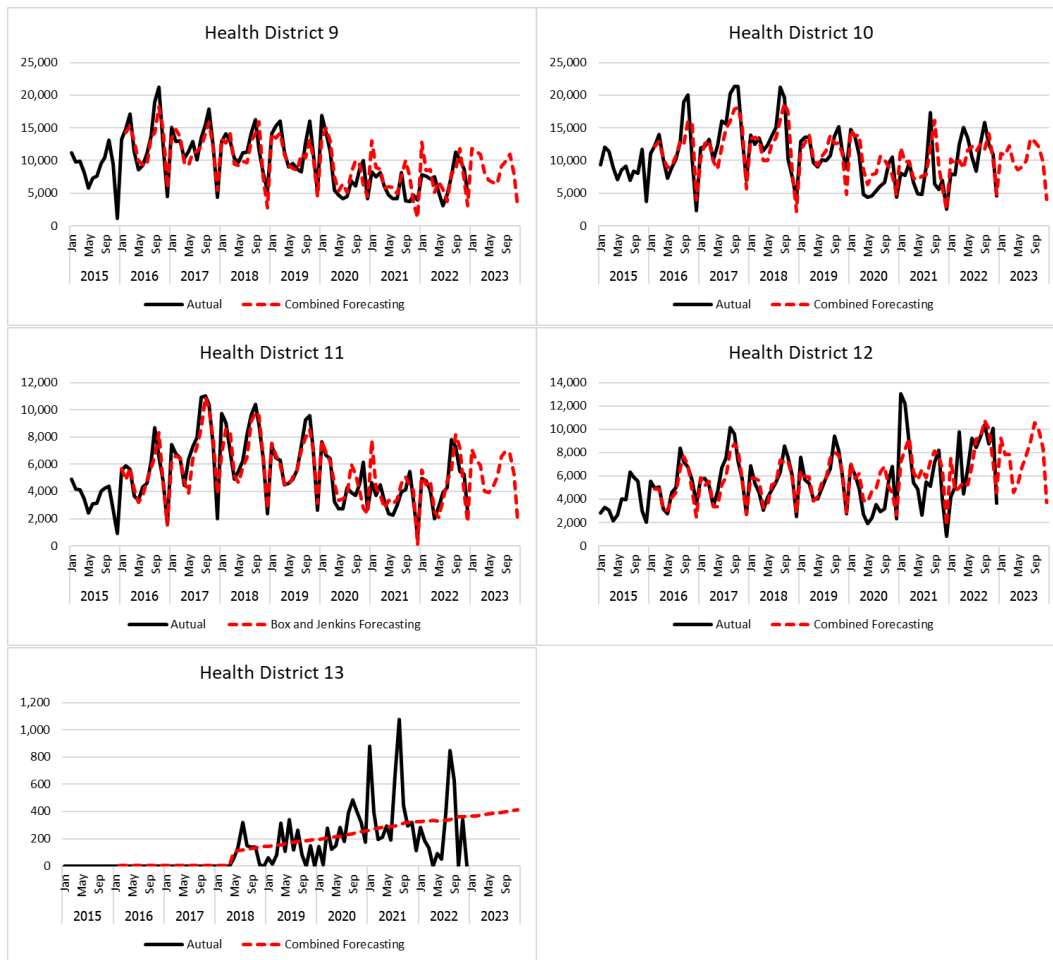
หมายเหตุ * ค่า sMAPE ต่ำที่สุด

หมายเหตุ	ค่าน้อย	ค่ามากสำหรับแต่ละเขตสุขภาพ
----------	---------	----------------------------

รูปที่ 2 การเปรียบเทียบค่าจริงกับค่าพยากรณ์โดยวิธีพยากรณ์ที่เหมาะสมเขตสุขภาพที่ 1 ถึง 8



รูปที่ 3 การเปรียบเทียบค่าจริงกับค่าพยากรณ์โดยวิธีพยากรณ์ที่เหมาะสมเขตสุขภาพที่ 9 ถึง 13



ตารางที่ 10 ร้อยละอิทธิพลฤดูกาลและร้อยละการเปลี่ยนแปลงของผู้ป่วยโรคปอดอักเสบปี พ.ศ. 2566 เมื่อเทียบกับปีก่อนแยกตามเขตสุขภาพ 13 เขต

Health District	2023												%change relative to the previous year
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
1	98	90	113	105	95	97	85	95	118	140	115	49	122
2	64	93	135	160	103	71	82	103	124	111	125	30	105
3	118	111	113	92	82	80	97	112	117	118	105	56	127
4	91	133	145	98	65	74	110	110	139	129	83	22	247
5	99	94	94	84	91	87	102	115	126	124	118	65	111
6	118	87	103	77	71	72	106	166	158	125	98	20	193
7	121	116	123	90	81	89	86	107	123	136	106	21	91
8	146	134	140	87	76	70	81	95	105	112	103	51	97
9	137	132	128	95	81	77	75	106	115	127	91	35	120
10	107	106	119	93	84	88	97	129	126	117	94	39	93
11	132	118	111	75	74	81	95	123	131	126	98	34	116
12	122	104	104	60	70	89	103	118	139	131	110	49	101
13	94	95	96	97	98	99	101	102	103	104	105	106	157

หมายเหตุ: ค่าร้อยละที่มีค่าเกิน 100

4. การอภิปราย

จากการพยากรณ์ด้วยตัวแบบที่เหมาะสมทั้ง 13 เขตสุขภาพของประเทศไทยตามตารางที่ 9 เมื่อนำผลค่าการพยากรณ์มาวิเคราะห์แนวโน้มและฤดูกาลได้ผลตาม ตารางที่ 10 ทำให้ทราบว่า เขตสุขภาพจำนวน 10 เขต คือ 1 2 3

4 5 6 9 11 12 และ 13 มีจำนวนผู้ป่วยเพิ่มขึ้นจากปี พ.ศ. 2565 โดยเขตที่ต้องเฝ้าระวังเพราะมีโอกาสผู้ป่วยเพิ่มขึ้นสูงที่สุดคือ เขตสุขภาพที่ 4 เพิ่มเป็นร้อยละ 247 รองลงมา เป็นเขตสุขภาพที่ 6 เพิ่มเป็นร้อยละ 193 ส่วนเขตสุขภาพที่ 7 8 และ 10 มีจำนวนผู้ป่วยลดลงจากปี พ.ศ. 2565

เล็กน้อย คือ ลดลงจากปกติร้อยละ 100 เหลือร้อยละ 91 97 และ 93 ตามลำดับเขตสุขภาพ

ในการพิจารณาฤดูกาลของโรค ทำให้ทราบความชุกของโรคปอดอักเสบว่าจะมีโอกาสเกิดขึ้นในช่วงไหนเป็นจำนวนมากเพื่อเพิ่มการเฝ้าระวังในพื้นที่ต่างๆ ได้อย่างมีประสิทธิภาพ พบว่าทุกเขตสุขภาพมีช่วงที่ต้องเฝ้าระวังเป็นพิเศษเพราะมีโอกาสเกิดโรคมามากที่สุดในช่วงฤดูฝนไปต้นฤดูหนาวคือประมาณช่วงเดือน สิงหาคม ถึงเดือนตุลาคม และช่วงฤดูหนาวไปต้นฤดูร้อนที่มีความเกิดการระบาดของโรคมามากกว่าปกติคือประมาณช่วงเดือน มกราคม ถึงเดือน มีนาคม ยกเว้นเขตสุขภาพที่ 1 2 5 และ 13 โดยในกรณีเขตสุขภาพที่ 1 และ 2 มีข้อสังเกตเพิ่มเติมเนื่องจากเป็นพื้นที่ตอนเหนือของประเทศไทยการระบาดของโรคมามากกว่าปกติจะเป็นช่วงเริ่มต้นฤดูร้อนคือช่วงเดือน มีนาคม ถึงเดือนเมษายน ทั้งนี้ควรมีการศึกษาเพิ่มเติมต่อไปว่าเหตุใดในเขตสุขภาพทั้ง 2 มีช่วงเวลากการป่วยโรคปอดอักเสบที่แตกต่างไปจากเขตสุขภาพอื่นๆ ซึ่งสมมติฐานเบื้องต้นอาจจะคาดการณ์จากภาวะฝุ่นควันที่เป็นปัญหาอย่างมากในทุกปีของพื้นที่ตอนเหนือของประเทศไทย นั้นอาจจะเป็นสาเหตุสำคัญที่ทำให้ผู้ป่วยโรคปอดอักเสบมีความชุกของโรคเพิ่มขึ้นเป็นพิเศษในช่วงฤดูร้อนต่างๆ ที่เขตสุขภาพอื่นๆ ช่วงเดือนเมษายนจะเป็นช่วงจำนวนผู้ป่วยลดลง

5. บทสรุป

การวิจัยครั้งนี้ได้นำเสนอการเลือกตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาข้อมูลผู้ป่วยโรคปอดอักเสบรายเดือนในเขตสุขภาพ 13 เขตของประเทศไทย โดยใช้อนุกรมเวลารายเดือนจากเว็บไซต์ของ Health Data Center กระทรวงสาธารณสุข พบว่า เขตสุขภาพจำนวน 11 เขต คือ 1 2 3 4 5 6 8 9 10 12 และ 13 ตัวแบบพยากรณ์ที่เหมาะสมคือ ตัวแบบพยากรณ์รวม โดยใช้วิธีถ่วงน้ำหนักด้วยสัมประสิทธิ์การถดถอย (REG) สอดคล้องกับงานวิจัย Keerativibool (2016) และ Nakunthod & Khamkhod (2018) ที่ศึกษาตัวแบบการพยากรณ์ผู้ป่วยโรคปอดอักเสบ อย่างไรก็ตามวิธีการพยากรณ์อื่นๆ ก็อาจจะเป็นวิธีที่ให้ความแม่นยำในการพยากรณ์ที่ดีได้เช่นเดียวกัน ขึ้นอยู่กับลักษณะของข้อมูลของเรื่องที่ทำการศึกษาและช่วงเวลาที่ทำการศึกษา ดังนั้นในการสร้างตัวแบบการพยากรณ์จึง

จำเป็นต้องทำการศึกษาวิธีการทางสถิติในหลายตัวแบบเพื่อใช้เป็นทางเลือกในการตัดสินใจที่ดีของงานวิจัย เช่น วิธีแยกส่วนประกอบที่มีส่วนประกอบของแนวโน้มและฤดูกาล รูปแบบคุณเป็นตัวแบบพยากรณ์ที่เหมาะสมของเขตสุขภาพที่ 7 และตัวแบบพยากรณ์บอซ-เจนกินส์เป็นตัวแบบพยากรณ์ที่เหมาะสมเขตสุขภาพที่ 11 เป็นต้น

ผลการศึกษาข้อมูลจริงและค่าพยากรณ์ที่ได้จากตัวแบบที่เหมาะสมในแต่ละเขตสุขภาพทั้ง 13 เขต พบว่า แนวโน้มผู้ป่วยโรคปอดอักเสบลดลงในช่วงปี พ.ศ. 2562 จนถึงปี พ.ศ. 2564 เกือบทุกเขตสุขภาพ ยกเว้นเขตสุขภาพที่ 4 6 12 และ 13 เท่านั้นที่เพิ่มขึ้น การลดลงของผู้ป่วยโรคปอดอักเสบนี้มีความเกี่ยวข้องโดยตรงกับช่วงสถานการณ์ระบาดของเชื้อโควิด-19 เนื่องจากประชากรไทยมีการระมัดระวังสุขภาพเกี่ยวกับโรคปอดกันมากขึ้นในช่วงเวลาปลายปี พ.ศ. 2562 จนถึงกลางปี พ.ศ. 2564 ที่เป็นช่วงการระบาดของโควิด-19 จึงเป็นผลดีต่อการลดจำนวนผู้ป่วยโรคปอดอักเสบของประเทศไทยในภาพรวมทั้งหมด อย่างไรก็ตามเมื่อฟื้นความตื่นตระหนกของวิกฤตการณ์ของโควิด-19 แล้ว คือช่วงหลังของปี พ.ศ. 2564 ถึง พ.ศ. 2565 ผู้ป่วยโรคปอดอักเสบมีจำนวนเพิ่มขึ้นเกือบทุกเขตสุขภาพ

โรคนี้ก็เป็นโรคประจำฤดูกาล ซึ่งจะเห็นได้ชัดจากตารางที่ 10 ว่าโรคนี้ส่วนใหญ่จะเป็นในช่วงปลายฤดูหนาวคาบเกี่ยวไปต้นฤดูร้อนและช่วงปลายฤดูฝนคาบเกี่ยวไปต้นฤดูหนาวของประเทศไทย ดังนั้นในช่วงเวลาดังกล่าวหน่วยงานสาธารณสุขของไทยยังจำเป็นต้องเฝ้าระวังการระบาดของโรคนี้เป็นพิเศษ

ในเขตสุขภาพที่ 13 เป็นเขตสุขภาพที่มีผลการพยากรณ์ที่ไม่แม่นยำ การใช้วิธีการพยากรณ์อนุกรมเวลาในการศึกษานี้อาจจะไม่สามารถจับรูปแบบจำนวนผู้ป่วยที่ถูกต้องของเขตสุขภาพนี้ได้ จำนวนผู้ป่วยโรคปอดอักเสบอาจจะขึ้นอยู่กับปัจจัยอื่นนอกเหนือจากเวลาหรือข้อมูลอนุกรมเวลา เช่น ระดับความเข้มข้นของค่า PM2.5 ในอากาศ อุณหภูมิ หรือปริมาณน้ำฝน เป็นต้น ดังนั้นการศึกษาดัชนีแบบถดถอยพหุคูณ (Multiple Regression) หรือ Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX) ทั้งสองตัวแบบนี้สามารถพิจารณาปัจจัยอื่นๆ นอกเหนือจากเวลาหรือข้อมูลอนุกรมเวลาในการพยากรณ์ได้ การศึกษาเพิ่มเติมอาจช่วยให้เข้าใจปัจจัยที่มีผลต่อจำนวนผู้ป่วยโรคปอด

อีกเสบนในเขตสุขภาพที่ 13 ได้ดียิ่งขึ้น และนำไปสู่การพัฒนาวิธีการพยากรณ์ที่แม่นยำยิ่งขึ้น

อย่างไรก็ตามวิธีการพยากรณ์อนุกรมเวลานั้นยังสามารถใช้เทคนิคอื่นๆ ทางด้านวิทยาการข้อมูลที่ใช้การเรียนรู้ของเครื่องช่วยในการหาคำตอบ เช่น การใช้โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network: RNN) (Hewamalage et al., 2021) เป็นเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ที่เป็นที่ยอมรับในปัจจุบัน เมื่อลองนำเทคนิคนี้มาใช้ในการพยากรณ์ผู้ป่วยโรคปอดอักเสบ โดยกำหนดค่าของไฮเปอร์พารามิเตอร์ ดังนี้ ฟังก์ชันความสูญเสีย (Loss Function) เป็นวิธีคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) ตัวเพิ่มประสิทธิภาพ (Optimizer) เป็นวิธี adam ทำการฝึกสอน (Epoch) จำนวน 100 ครั้ง และจำนวนข้อมูลย้อนหลัง (Look Back) เท่ากับ 12 และ 24 ประมวลผลด้วยไพทอน (Python) แล้วเลือกค่า sMAPE ที่ต่ำที่สุดของแต่ละเขตสุขภาพจากการทดลอง ได้ค่าร้อยละ sMAPE ดังตารางที่ 11

ตารางที่ 11 ค่า sMAPE ของ RNN แยกตามเขตสุขภาพ 13 เขต ในชุดข้อมูล

ทดสอบ

Health District	RNN
1	35.12
2	82.07
3	43.07
4	141.97
5	38.18
6	90.39
7	117.60
8	127.48
9	40.14
10	40.03
11	46.33
12	37.53
13	99.81

ค่าร้อยละของ sMAPE ของทั้ง 13 เขตสุขภาพจากตารางที่ 11 นั้น บางเขตสุขภาพได้ผลลัพธ์ที่ดีแต่ยังได้ผลไม่ดีกว่าค่าต่ำที่สุดจากวิธีทางสถิติในตารางที่ 8 อย่างไรก็ตามผู้ที่สนใจศึกษาทางด้านเทคนิคการเรียนรู้ของเครื่องอาจจะศึกษาเชิงลึกและทำการทดลองเพิ่มเติมเรื่องการกำหนดค่าของไฮเปอร์พารามิเตอร์ที่แตกต่างจากงานวิจัยนี้เพื่อให้ค่าพยากรณ์แม่นยำมากขึ้นได้

เอกสารอ้างอิง

- Armstrong, J. S. (1985). *Long-range Forecasting: From Crystal Ball to Computer*, 2nd ed. Wiley.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed. New Jersey: Prentice Hall.
- Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health. (2015). *Annual Epidemiological Surveillance Report 2015*. Bangkok, Thailand: The Agricultural Co-operative Federation of Thailand., Ltd. Publisher. Retrieved May 3, 2023, from <https://apps-doe.moph.go.th/boceng/annual/Annual/AESR2015/> (in Thai)
- Charupronprasit, R., Sirirat, B., & Sumtip, J. (2017). *Statistical Thailand 2017*. Nontaburi, Bureau of Policy and Strategy, Ministry of Public Health. Retrieved May 3, 2023, from <http://digital.nlt.go.th/dlib/items/show/10642> (in Thai)
- Charupronprasit, R., Suthart, A., Sirirat, B., & Dumrongchat, B. (2016). *Statistical Thailand 2016*. Nontaburi, Bureau of Policy and Strategy, Ministry of Public Health. Retrieved May 3, 2023, from <http://digital.nlt.go.th/dlib/items/show/10640> (in Thai)
- Ekachampaka, P., Sriprayoon, P., & Thammachart, S. (2018). *Thailand Regional Health Profile 2012 - 2017*. Strategy and Planning Division of Office of the Permanent Secretary Ministry of Public Health, Bangkok: Chinauksorn Printing. (In Thai)
- Flores, B. E. (1986). A pragmatic view of accuracy measurement in forecasting. *Omega*, 14(2), 93-98.
- Health Data Center, Ministry of Public Health. (2023). *Rates of Pneumonia Cases*. Retrieved May 3, 2023, from <https://hdcservice.moph.go.th/hdc/main/index.php>
- Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388-427.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.
- Keerativibool, W. (2016). Forecasting model for the number of patients with pneumonia in Thailand. *Public Health Journal Burapha University*, 11(1), 24-38. (in Thai).
- Minsan, P. (2022). Forecasting model for export condom quantity of Thailand on the COVID-19 situation. *The Journal of Applied Science*, 21(1), 1-21. (in Thai). <https://ph01.tci-thaijo.org/index.php/JASCI/article/view/244500>
- Minsan, W., & Minsan, P. (2023). Incorporating decomposition and the holt-winters method into the whale optimization algorithm for forecasting monthly government revenue in Thailand. *Science & Technology Asia*, 28(4), 38-53. <https://ph02.tci-thaijo.org/index.php/SciTechAsia/article/view/250335>
- Minsan, W., Saengngammuang, N., Taninpong, P., & Thumronglaohapun, S. (2021). Comparing methods of optimization in solver of excel 2019 and whale optimization algorithm. *UTK Journal*, 15(2), 107-120. (in Thai). <https://ph02.tci-thaijo.org/index.php/rmutk/article/view/245178>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to Linear Regression Analysis*, 4th ed. New York: Wiley.
- Nakunthod, I., & Khamkhod, K. (2018). Time series modeling for the pneumonia rate of patients in Lampang province. In: *Proceeding of the 18th Graduate Studies of Northern Rajabhat, University Network Conference and the 4th Lampang Research*, 20 July 2018, pp. 340-358. Lampang, Thailand. (in Thai)
- Persons, W. (1919). Indices of business conditions. *Review of Economics and Statistics*, 1, 5-110.

- Saeying, J., Minsan, W., & Taninpong, P. (2023). Forecasting model for the amount of water flowing into the reservoirs of the electricity generating authority of Thailand (EGAT). *RMUTSV Research Journal*, 15(2), 494-510. (in Thai). <https://li01.tci-thaijo.org/index.php/rmutsvrj/article/view/252100>
- Taesombat, S. (2006). *Quantitative Forecasting*. Bangkok: Kasetsart University Press. (in Thai)
- Tunkaew, T., Minsan, P., Nontapa, C., & Minsan, W. (2023). A suitable forecasting model for exchange rates of the top 10 foreign currencies most preferred by Thai tourists compared to the Thai baht. *Thai Science and Technology Journal*, 31(3), 1-21. (in Thai) <https://li01.tci-thaijo.org/index.php/tstj/article/view/258445>