

Research Article

การพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย

The Statistical Forecasting of COVID-19 Infected People in Thailand

จุฑารัตน์ วรประทีป *

Jutharath Voraprateep *

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง เขตบางกะปิ กรุงเทพมหานคร 10240

Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bang Kapi, Bangkok 10240, Thailand.

ABSTRACT

Article history:

Received: 2023-10-30

Revised: 2024-01-22

Accepted: 2024-02-27

Keywords:

COVID-19;

global constant mean model;

local constant mean model;

adaptive forecasting;

Box – Jenkins

The research objectives are to compare the statistical forecasting of COVID-19 infected people in Thailand using four techniques: global constant mean model, local constant mean model, adaptive forecasting and Box - Jenkins; and to study the errors from the statistical forecasting using MAD, MSE and MAPE to select the appropriate forecasting. This research has taken the number of COVID-19 infected people from the Center for COVID-19 Situation Administration (CCSA) between 1st April 2022 and 31st October 2022 (7 months or 214 days) for forecasting COVID-19 infected people in Thailand. The research results were summarized as follows: the minimum number of COVID-19 infected people was 26 persons in April and the maximum was 23,418 in August. The average number of COVID-19 infected people for 7 months is 8,802 persons, and the top three with the least MAD, MSE and MAPE for statistical forecasting of COVID-19 infected people in Thailand are single exponential smoothing with smoothing constants of $\alpha = 0.1$ and 0.5 , and single moving average with $n = 2$.

© 2025 Voraprateep, J. Recent Science and Technology published by Rajamangala University of Technology Srivijaya

บทคัดย่อ

การวิจัยเรื่อง การพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย มีวัตถุประสงค์การวิจัยเพื่อเปรียบเทียบวิธีการพยากรณ์ 4 เทคนิค ได้แก่ ตัวแบบค่าเฉลี่ยคงที่ระยะยาว ตัวแบบค่าเฉลี่ยคงที่ระยะสั้นการพยากรณ์แบบปรับได้ และ Box – Jenkins และทำการศึกษาความคลาดเคลื่อนที่เกิดจากการพยากรณ์โดยใช้เกณฑ์ MAD MSE และ MAPE ในการเลือกวิธีการพยากรณ์ที่เหมาะสม การวิจัยในครั้งนี้ใช้ข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 จากการรายงานของศูนย์บริหารสถานการณ์โควิด-19 (ศบค.) ระหว่างวันที่ 1 เมษายน 2564 - 31 ตุลาคม 2564 เป็นเวลา 7 เดือน (214 วัน) เพื่อพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย ผลการวิจัยพบว่าจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยมีค่าต่ำสุดในเดือนเมษายน เท่ากับ 26 คน และมีค่าสูงสุดในเดือนสิงหาคม เท่ากับ 23,418 คน โดยจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเฉลี่ย 7 เดือน เท่ากับ 8,802 คนการพยากรณ์ทางสถิติจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยที่มีค่า MAD,

* Corresponding author.

E-mail address: jutharath@ru.ac.th

Cite this article as:

Voraprateep, J. 2025. The Statistical Forecasting of COVID-19 Infected People in Thailand. *Recent Science and Technology* 17(1): 261074.

MSE และ MAPE น้อยที่สุด 3 อันดับแรก ได้แก่ การทำให้เรียบแบบเอกซ์โพเนนเชียลขั้นเดียว เมื่อค่าคงที่การทำให้เรียบ (a) เท่ากับ 0.1, 0.5 และวิธีค่าเฉลี่ยเคลื่อนที่ขั้นเดียว เมื่อจำนวนเทอม (n) เท่ากับ 2

คำสำคัญ: โควิด-19, ตัวแบบค่าเฉลี่ยคงที่ระยะยาว, ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น, การพยากรณ์แบบปรับได้, Box – Jenkins

1. บทนำ

ไวรัสโคโรนา (Coronavirus) เป็นไวรัสที่ถูกพบครั้งแรกในปี 1960 แต่ยังไม่ทราบแหล่งที่มาอย่างชัดเจนว่ามาจากที่ใด แต่เป็นไวรัสที่สามารถติดเชื่อได้ทั้งในมนุษย์และสัตว์ ปัจจุบันมีการค้นพบไวรัสสายพันธุ์นี้แล้วทั้งหมด 6 สายพันธุ์ ส่วนสายพันธุ์ที่กำลังแพร่ระบาดหนักทั่วโลกตอนนี้เป็นสายพันธุ์ที่ยังไม่เคยพบมาก่อน คือ สายพันธุ์ที่ 7 จึงถูกเรียกว่าเป็น “ไวรัสโคโรนาสายพันธุ์ใหม่” และในภายหลังถูกตั้งชื่ออย่างเป็นทางการว่า “โควิด-19” (COVID-19) นั้นเอง การแพร่ระบาดของไวรัสชนิดนี้ถูกพบครั้งแรกที่เมืองอู่ฮั่น เมืองเอกของมณฑลหูเป่ย์ ประเทศจีน ในช่วงปลายปี 2019 และ การแพร่ระบาดได้เกิดขึ้นอย่างต่อเนื่องและรุนแรงในหลายประเทศทั่วโลก ซึ่งในปัจจุบันยังไม่มียาหรือวัคซีนเพื่อใช้รักษาผู้ป่วยที่ติดเชื่อไวรัสโคโรนาสายพันธุ์ใหม่ หรือไวรัสโควิด-19 ได้ (Faculty of Medicine Siriraj Hospital, 2021)

สถานการณ์การแพร่ระบาดของเชื้อโควิด-19 ในประเทศไทยนั้น เริ่มมีการแพร่ระบาดประมาณต้นเดือนมีนาคม 2563 ซึ่งผู้ติดเชื้อมีความเกี่ยวข้องกับกลุ่มผู้ติดเชื้อจากสถานบันเทิงและสนามมวย นอกจากนี้ยังมีความเชื่อมโยงกับชาวต่างชาติและชาวไทยที่เดินทางกลับจากต่างประเทศหรือผู้ที่ทำงานกับนักท่องเที่ยวจากต่างประเทศ นับเป็นการแพร่ระบาดของเชื้อโควิด-19 ในประเทศไทยรอบที่ 1 (ประมาณเดือนมีนาคม – พฤษภาคม 2563) โดยหน่วยงานด้านสาธารณสุขของประเทศไทยและหน่วยงานที่เกี่ยวข้องทุกภาคส่วนได้ร่วมมือ ร่วมแรงและร่วมใจ ควบคุมการแพร่ระบาดของเชื้อไวรัสโควิด-19 ได้เป็นอย่างดี หลังจากนั้นการแพร่ระบาดของเชื้อไวรัสโควิด-19 ได้กลับมาแพร่ระบาดอีกครั้งจากกลุ่มตลาดกลางการค้าชายฝั่งที่จังหวัดสมุทรสาครในเดือนมกราคม 2564 เป็นรอบที่ 2 และได้กระจายการแพร่ระบาดออกไปยังกลุ่มต่าง ๆ ที่เกี่ยวข้อง ต่อมาประมาณต้นเดือนเมษายน 2564 ได้เกิดการแพร่ระบาดของเชื้อไวรัสโควิด-19 เป็นรอบที่ 3 จากสถานบันเทิงย่านทองหล่อ กรุงเทพฯ ซึ่งเป็นรอบที่มีการแพร่ระบาดที่รุนแรงมีผู้ติดเชื้อและเสียชีวิตเป็นจำนวนมาก (Ministry of Public Health, 2021)

การศึกษาทางวิชาการที่เกี่ยวข้องกับสถานการณ์การแพร่ระบาดของเชื้อไวรัสโควิด-19 มีหลากหลายมิติหนึ่งในการศึกษาที่ได้รับความสนใจคือ การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ซึ่งงานวิจัยที่พยากรณ์

จำนวน ผู้ติดเชื้อไวรัสโควิด-19 มีความแตกต่างกันในวิธีการศึกษาและผลสรุปที่ได้ เช่น การใช้จำนวนผู้ติดเชื้อสะสม ผู้เสียชีวิตและรักษาหายในประเทศไทยเพื่อศึกษาการใช้แบบจำลองคณิตศาสตร์ในการพยากรณ์ การศึกษาข้อมูลผู้ป่วยโดยใช้ข้อมูลประเทศอื่น ๆ มาใช้การพยากรณ์ตัวอย่างการศึกษาวิจัย ได้แก่ การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโคโรนา 2019 สะสมโดยใช้การเรียนรู้เชิงลึก (Nuanchui, 2021) การใช้แบบจำลองคณิตศาสตร์คาดการณ์การระบาดของโรคติดเชื้อไวรัสโควิด-19 (Auswawanlop and Aeiutrakul, 2021) แบบจำลองการพยากรณ์จำนวนผู้ป่วยโรคติดเชื้อไวรัสโคโรนา 2019 สะสมที่มีผู้เสียชีวิตและรักษาหายในประเทศไทย (Areepong and Sunthornwat, 2021) การวิเคราะห์เชิงคาดการณ์ของผู้ป่วยโรคโควิด-19 ในประเทศไทยโดยใช้ข้อมูลหลายประเทศ (Vorathamthongdee and Congstitvatana, 2023) การพยากรณ์การแพร่ระบาดของไวรัสโควิด-19 ในประเทศไทยแบบเรียลไทม์โดยใช้วิธีการแบบไม่ใช้แบบจำลองอย่างง่ายและแบบจำลองการถดถอยอนุกรมเวลา (Wongsathan, 2021) และการพยากรณ์ยอดผู้ติดเชื้อ COVID-19 โดยใช้หลักทางสถิติ (Panichayakul, 2020)

จะเห็นได้ว่างานวิจัยในอดีตที่เกี่ยวข้องกับการพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเป็นการวิจัยที่ใช้วิธีการหรือตัวแบบทางคณิตศาสตร์ หรือตัวแบบอื่น ๆ ทางระบาดวิทยาในการศึกษา โดยใช้ข้อมูลจำนวนผู้ติดเชื้อในช่วงเวลาและข้อจำกัดที่แตกต่างกันสำหรับการพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเรื่องนี้ ได้นำข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 รายวัน ซึ่งถูกนำเสนอเป็นประจำวันโดยศูนย์บริหารสถานการณ์โควิด-19 (ศบค.) มาใช้ในการศึกษาด้วยวิธีการพยากรณ์ 4 เทคนิค ได้แก่ ตัวแบบค่าเฉลี่ยคงที่ระยะยาว ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น การพยากรณ์แบบปรับได้ และ Box – Jenkins และทำการศึกษาความคลาดเคลื่อนที่เกิดจากการพยากรณ์โดยใช้เกณฑ์ MAD MSE และ MAPE ในการเลือกวิธีการพยากรณ์ที่เหมาะสม ซึ่งผลการศึกษาจะทำให้ทราบวิธีการพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยที่เหมาะสมที่สุด โดยผลการศึกษาจะเป็นประโยชน์ต่อการการศึกษาด้านสถิติและหน่วยงานที่เกี่ยวข้อง นอกจากนี้ผลการศึกษานำมาใช้ในการประกอบการเรียนการสอนให้กับนักศึกษาในทุกระดับชั้น

2. วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัย มีรายละเอียดดังต่อไปนี้

2.1 การจัดเตรียมข้อมูล

ข้อมูลที่นำมาใช้ในการวิจัย คือ ข้อมูลทุติยภูมิ ซึ่งเป็นข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 จากการรายงานของศูนย์บริหารสถานการณ์โควิด-19 (ศบค.) ระหว่างวันที่ 1 เมษายน 2564 - 31 ตุลาคม 2564 เป็นเวลา 7 เดือน หรือ 214 วัน โดยข้อมูลที่ใช้ในการพยากรณ์ครั้งนี้จัดเป็นข้อมูลอนุกรมเวลาที่มีการจัดเก็บอย่างต่อเนื่องทุกวัน สำหรับข้อมูลในช่วงเวลา 7 เดือนดังกล่าว เป็นช่วงเวลาการแพร่ระบาดของเชื้อไวรัสโควิด-19 รอบที่ 3 ของประเทศไทย การเลือกใช้ข้อมูลในช่วงเวลานี้มาทำการศึกษา เนื่องจากเป็นช่วงเวลาที่การแพร่ระบาดมีความรุนแรงมากที่สุดของประเทศไทย กล่าวคือ มีจำนวนผู้ติดเชื้อไวรัสโควิด-19 และจำนวนผู้เสียชีวิตเป็นจำนวนมาก มาตรการต่าง ๆ ที่ภาครัฐนำมาใช้ในการควบคุมการแพร่ระบาดมีความเข้มงวดเป็นอย่างมาก ดังนั้น ข้อมูลจำนวนผู้ติดเชื้อ ในช่วงเวลาดังกล่าว จึงมีความสมบูรณ์และน่าสนใจในการนำมาใช้ศึกษาการพยากรณ์ทางสถิติ

2.2 การศึกษาความเคลื่อนไหวของข้อมูล

การศึกษาค่าเคลื่อนไหวของข้อมูลอนุกรมเวลาที่ศึกษา เป็นการพิจารณาเบื้องต้นว่าข้อมูลอนุกรมเวลาจำนวนผู้ติดเชื้อไวรัสโควิด-19 มีลักษณะแบบใดหรือมีแนวโน้มของข้อมูลเป็นอย่างไร โดยการพิจารณาจากแผนภาพที่แสดงความสัมพันธ์ระหว่างเวลาหน่วยเป็นวัน (t) และจำนวนผู้ติดเชื้อไวรัสโควิด-19 (X_t) ด้วยโปรแกรม Microsoft Excel 365

2.3 การวิเคราะห์ข้อมูลทางสถิติ

การวิเคราะห์ข้อมูลทางสถิติและการสร้างแผนภาพในการวิจัยนี้ได้ใช้โปรแกรม Microsoft Excel 365 และโปรแกรม R สำหรับการพยากรณ์ทางสถิติหรือการพยากรณ์เชิงปริมาณที่ใช้มีรายละเอียดพอสังเขป ดังนี้ (Lorjirashunkul and Jittawej, 2005)

2.3.1 ตัวแบบค่าเฉลี่ยคงที่ระยะยาว (global constant mean model) ตัวแบบนี้เขียนแทนด้วยสมการดังนี้

$$X_t = \mu + \varepsilon_t \quad , \quad t = 1, 2, 3, \dots \quad (1)$$

เมื่อ X_t คือ ค่าของตัวแปร X ณ เวลา t

μ คือ ค่าเฉลี่ยของข้อมูล (ค่าคงที่)

ε_t คือ ความคลาดเคลื่อนสุ่ม ณ เวลา t ซึ่ง ε_t เป็นอิสระกันในแต่ละเวลา t ใด ๆ นั่นคือ ε_i และ ε_j เป็นอิสระกัน เมื่อ $i \neq j$ โดยมีข้อตกลงเบื้องต้นเกี่ยวกับ ε_t คือ ε_t มีค่าคาดหวังเท่ากับศูนย์และมีความแปรปรวนคงที่เท่ากับ σ^2

ดังนั้น ค่าพยากรณ์ k หน่วยเวลาล่วงหน้า เมื่อพยากรณ์ ณ เวลา t หรือ X_{t+k} เขียนแทนด้วย $\hat{X}_t(k)$ นั่นคือค่าพยากรณ์ X_{t+k} ซึ่งพยากรณ์ ณ เวลา t คือค่าคาดหวังของสมการ (1) จึงเขียนได้เป็น

$$\hat{X}_t(k) = \mu \quad (2)$$

จากสมการ (2) ค่า μ คือ พารามิเตอร์ที่เราไม่ทราบค่าที่แท้จริง ดังนั้นเราประมาณค่า μ ด้วยค่าเฉลี่ยของข้อมูลที่มีอยู่ทั้งหมดในขณะนั้น ดังนี้

$$\hat{X}_t(k) = \frac{\sum_{i=1}^t X_i}{t} \quad , \quad k = 1, 2, 3, \dots \quad (3)$$

2.3.2 ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น (local constant mean model) ตัวแบบนี้เขียนแทนด้วยสมการดังนี้

$$X_t = \mu_t + \varepsilon_t \quad , \quad t = 1, 2, 3, \dots \quad (4)$$

เมื่อ X_t คือ ค่าของตัวแปร X ณ เวลา t

μ_t คือ ค่าเฉลี่ยของข้อมูล (ค่าคงที่) ณ เวลา t (ช่วงเวลาสั้น ๆ)

ε_t คือ ความคลาดเคลื่อนสุ่ม ณ เวลา t ซึ่ง ε_t เป็นอิสระกันในแต่ละเวลา t ใด ๆ นั่นคือ ε_i และ ε_j เป็นอิสระกัน เมื่อ $i \neq j$ โดยมีข้อตกลงเบื้องต้นเกี่ยวกับ ε_t คือ ε_t มีค่าคาดหวังเท่ากับศูนย์และมีความแปรปรวนคงที่เท่ากับ σ^2

สำหรับวิธีการพยากรณ์ค่า X_t ภายใต้ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น ได้แก่

(1) วิธีค่าเฉลี่ยเคลื่อนที่ขั้นเดียว (single moving average) ถ้าจำนวนข้อมูลในช่วงเวลาสั้น ๆ นั้นมีอยู่ n เทอม แล้วค่าพยากรณ์หนึ่งหน่วยเวลาล่วงหน้าที่พยากรณ์ ณ เวลา t จึงเท่ากับค่าเฉลี่ยเลขคณิตของข้อมูลจำนวน n เทอม ดังนี้

$$\hat{X}_t(1) = \frac{X_{t-n+1} + X_{t-n+2} + \dots + X_t}{n} \quad (5)$$

ข้อมูล	X_1	X_2	X_3	X_4	...	X_{t-2}	X_{t-1}	X_t
น้ำหนัก	a^{t-1}	a^{t-2}	a^{t-3}	a^{t-4}	...	a^2	a^1	a^0

เมื่อ a คือ ค่าคงที่ เรียกว่า ค่าคงที่การทำให้เรียบ (smoothing constant) และ $0 < a < 1$ ค่าพยากรณ์หนึ่งหน่วยเวลาล่วงหน้า เมื่อพยากรณ์ ณ เวลา t ของวิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลขั้นเดียว จึงเป็นค่าเฉลี่ยถ่วงน้ำหนัก ดังนี้

$$\hat{X}_t(1) = \frac{\sum_{r=0}^{t-1} a^r X_{t-r}}{\sum_{r=0}^{t-1} a^r} \quad (6)$$

2.3.3 การพยากรณ์แบบปรับได้ โดยวิธีการพยากรณ์แบบการกรองที่ปรับได้ อาศัยแนวคิดที่จะสร้างตัวแบบ การพยากรณ์ที่สามารถปรับค่าพารามิเตอร์ได้ (adaptive forecasting model) สำหรับการวิจัยนี้เลือกใช้การพยากรณ์แบบการกรองที่ปรับได้ (adaptive filtering) ค่าพยากรณ์จะเขียนได้ในเทอมของข้อมูลในอดีต n ตัว ได้ดังนี้

$$\hat{X}_t(1) = W_0 X_t + W_1 X_{t-1} + \dots + W_{n-1} X_{t-n+1} \quad (7)$$

โดยที่สัมประสิทธิ์ W_i ไม่เป็นค่าคงที่ แต่สามารถปรับค่าให้สอดคล้องกับการเปลี่ยนแปลงของข้อมูลได้

2.3.4 Box - Jenkins เป็นวิธีวิเคราะห์ข้อมูลอนุกรมเวลาในกรณีที่ข้อมูลมีความเกี่ยวข้องหรือสัมพันธ์กัน โดยทำการศึกษาจากฟังก์ชันสหสัมพันธ์ในตัวเอง (autocorrelation function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (partial autocorrelation function: PACF) Box - Jenkins ได้จำแนกตัวแบบการพยากรณ์ออกเป็น 4 ตัวแบบคือ

(1) ตัวแบบการเฉลี่ยเคลื่อนที่หรือตัวแบบเอ็มเอ (Moving Average Model: MA model)

ตัวแบบการเฉลี่ยเคลื่อนที่อันดับ q หรือ MA(q) เขียนได้ดังนี้

(2) การทำให้เรียบแบบเอกซ์โพเนนเชียลขั้นเดียว (single exponential smoothing) ให้น้ำหนักของข้อมูลแตกต่างกันที่หน่วยเวลาต่าง ๆ โดยจะให้น้ำหนักของข้อมูลลดลงแบบเอกซ์โพเนนเชียลตามหน่วยเวลาที่ห่างออกไปในอดีต ลักษณะการให้น้ำหนักกับข้อมูลเป็นดังนี้

$$X_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (8)$$

เมื่อ X_t คือ ค่าของตัวแปร X ณ เวลา t

μ คือ ค่าเฉลี่ยของข้อมูล (ค่าคงที่)

$\theta_1, \theta_2, \dots, \theta_q$ คือ พารามิเตอร์ (ค่าคงที่) ของตัวแบบ

$u_t, u_{t-1}, \dots, u_{t-q}$ คือ ความคลาดเคลื่อนสุ่ม (white noise or error terms) ซึ่งเป็นอิสระกันและมีการแจกแจงความน่าจะเป็นแบบเดียวกัน มีค่าคาดหวังเท่ากับศูนย์ และมีความแปรปรวนคงที่เท่ากับ σ^2

(2) ตัวแบบการถดถอยในตัวเองหรือตัวแบบเออาร์ (Autoregressive Model: AR model)

ตัวแบบการถดถอยในตัวเองอันดับ p หรือ AR(p) เขียนได้ดังนี้

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \delta + u_t \quad (9)$$

เมื่อ X_t คือ ค่าของตัวแปร X ณ เวลา t

δ คือ ค่าคงที่

$\phi_1, \phi_2, \dots, \phi_p$ คือ พารามิเตอร์ (ค่าคงที่) ของตัวแบบ

u_t คือ ความคลาดเคลื่อนสุ่ม (white noise or error terms)

(3) ตัวแบบผสมการถดถอยในตัวเองกับการเฉลี่ยเคลื่อนที่หรือตัวแบบอาร์มา (Mixed Autoregressive Moving Average Model: ARMA model)

ตัวแบบผสมการถดถอยในตัวเองกับการเฉลี่ยเคลื่อนที่อันดับ (p, q) หรือ ARMA(p, q) คือตัวแบบที่มีเทอมของตัวแบบการถดถอยในตัวเอง (AR) อยู่ p เทอมและตัวแบบการเฉลี่ยเคลื่อนที่ (MA) อยู่ q เทอม และเขียนได้ดังนี้

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \delta + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (10)$$

(4) ตัวแบบรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่หรือตัวแบบอาร์มา (Autoregressive Integrated Moving Average Model: ARIMA model) เป็นตัวแบบที่นำไปประยุกต์ใช้กับข้อมูลอนุกรมเวลาที่ไม่คงที่ (nonstationary)

กำหนดให้ W_t คือผลต่างที่หนึ่ง (first difference) ของข้อมูลอนุกรมเวลา ดังนั้น

$$W_t = X_t - X_{t-1} \quad (11)$$

ตัวแบบทั่วไปสำหรับ W_t เขียนได้เป็น

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (12)$$

ดังนั้น ตัวแบบสำหรับ X_t จึงเป็น

$$X_t = X_{t-1} + \phi_1 (X_{t-1} - X_{t-2}) + \phi_2 (X_{t-2} - X_{t-3}) + \dots + \phi_p (X_{t-p} - X_{t-p+1}) + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q} \quad (13)$$

แต่ X_t เขียนอยู่ในรูปผลบวกของ W_t ได้เป็น

$$X_t = W_1 + W_{t-1} + W_{t-2} + \dots \quad (14)$$

สมการ (12) ถูกเรียกว่า ตัวแบบรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่หรือตัวแบบอาร์มา (Autoregressive Integrated Moving Average Model: ARIMA model)

การพยากรณ์โดยวิธี Box - Jenkins มีขั้นตอนในการวิเคราะห์ดังนี้

1) การตรวจสอบข้อมูลอนุกรมเวลาเพื่อพิจารณาว่าเป็นอนุกรมเวลาคงที่ (stationary) หรือไม่ โดยพิจารณาจากกราฟฟังก์ชันสหสัมพันธ์ในตัว (autocorrelation function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวบางส่วน (partial autocorrelation function: PACF) ของข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 (X_t)

2) ในกรณีที่ข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 เป็นข้อมูลอนุกรมเวลาไม่คงที่ พิจารณาการเลือกใช้ผลต่างที่หนึ่ง, สอง, สาม, ... ทั้งนี้ จะใช้ผลต่างของข้อมูลอนุกรมเวลาอนุกรมแรกที่สุดที่เป็นข้อมูลอนุกรมเวลาที่จะพิจารณาต่อไป โดยพิจารณาจากกราฟ ACF และ PACF ประกอบการตัดสินใจว่าผลต่างของข้อมูลอนุกรมเวลาคงที่หรือไม่

3) การกำหนดตัวแบบ เป็นการหาตัวแบบอนุกรมเวลาที่คิดว่าเหมาะสมกับข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19

4) การประมาณค่าพารามิเตอร์ของตัวแบบที่คิดว่าเหมาะสม

5) การตรวจสอบความถูกต้องและเหมาะสมของตัวแบบโดยการทดสอบคุณสมบัติความสุ่มของค่าคลาดเคลื่อนสุ่ม u_t ด้วยตัวสถิติทดสอบ Ljung - Box และการตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ด้วยเกณฑ์สารสนเทศของอะกะอิเกะหรือเอไอซี (Akaike's Information Criterion: AIC)

1. การทดสอบคุณสมบัติความสุ่มของค่าคลาดเคลื่อนสุ่ม u_t ด้วยตัวสถิติทดสอบ Box - Pierce หรือตัวสถิติทดสอบ Ljung - Box โดยมีสมมุติฐานในการทดสอบดังนี้

$H_0: u_t$ มีคุณสมบัติความสุ่ม

$H_a: u_t$ ไม่มีคุณสมบัติความสุ่ม

ถ้าสมมุติฐานว่าง (H_0) ไม่ได้รับการปฏิเสธ แล้ว u_t มีคุณสมบัติความสุ่ม

2. การตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ด้วยเกณฑ์สารสนเทศของอะกะอิเกะหรือเอไอซี (Akaike's Information Criterion: AIC) ซึ่งเป็นเกณฑ์การคัดเลือกตัวแบบที่เหมาะสมภายใต้ตัวแบบที่มีการใช้ข้อมูลชุดเดียวกันแต่มีจำนวนพารามิเตอร์แตกต่างกัน ตัวแบบที่เหมาะสมกว่า คือ ตัวแบบที่มีค่าเอไอซีต่ำกว่านั่นเอง

2.4 การเปรียบเทียบวิธีการพยากรณ์

การเปรียบเทียบวิธีการพยากรณ์ ทำได้โดยการคำนวณค่าพยากรณ์ของจำนวนผู้ติดเชื้อไวรัสโควิด-19 จากวิธีการพยากรณ์ในตัวแบบต่าง ๆ และคำนวณค่าความคลาดเคลื่อนจากการพยากรณ์ ได้แก่ (Sinsomboonthong, 2016)

3.2 การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย

การพยากรณ์เริ่มต้นจากการศึกษาความเคลื่อนไหวของข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย เพื่อเป็นการพิจารณาเบื้องต้นว่ามีลักษณะแบบใดหรือมีแนวโน้มของข้อมูลเป็นอย่างไร จากการสร้างแผนภาพที่แสดงความสัมพันธ์ระหว่างเวลา หน่วยเป็นวัน (t) และจำนวนผู้ติดเชื้อไวรัสโควิด-19 (X_t) แสดงดังภาพที่ 1 พบว่าจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเพิ่มสูงขึ้นจากวันที่ 1 เมษายน 2564 ซึ่งมีจำนวนผู้ติดเชื้อไวรัส

โควิด-19 ต่ำสุดจำนวน 26 คน (ลำดับวันที่ 1) จนถึงวันที่ 13 สิงหาคม 2564 ซึ่งมีจำนวนผู้ติดเชื้อไวรัสโควิด-19 สูงสุดจำนวน 23,418 คน (ลำดับวันที่ 135) หลังจากนั้นจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเริ่มลดลงตามลำดับ

หลังจากศึกษาความเคลื่อนไหวของข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเรียบร้อยแล้ว จึงเริ่มดำเนินการพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย ซึ่งผลการศึกษาโดยใช้การพยากรณ์ทางสถิติตัวแบบต่าง ๆ เป็นดังนี้

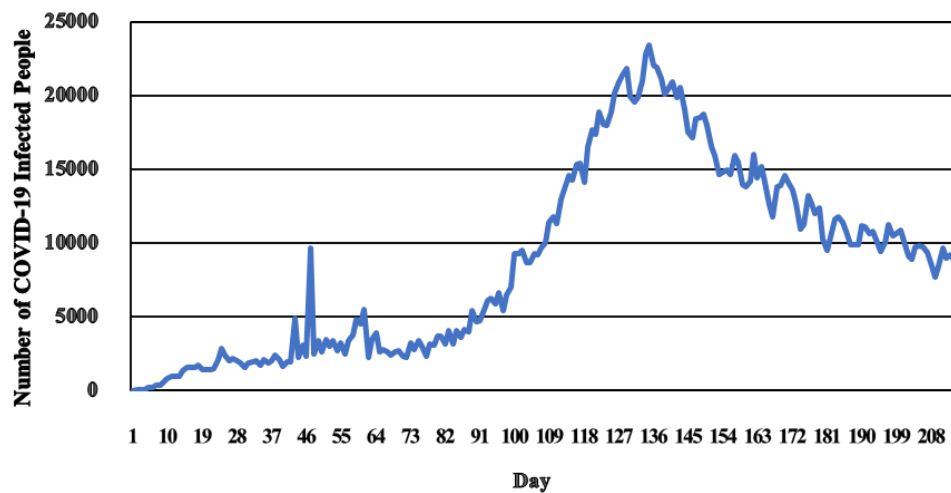


Figure 1 Line chart for COVID-19 infected people in Thailand between April 1, 2021 – October 31, 2021

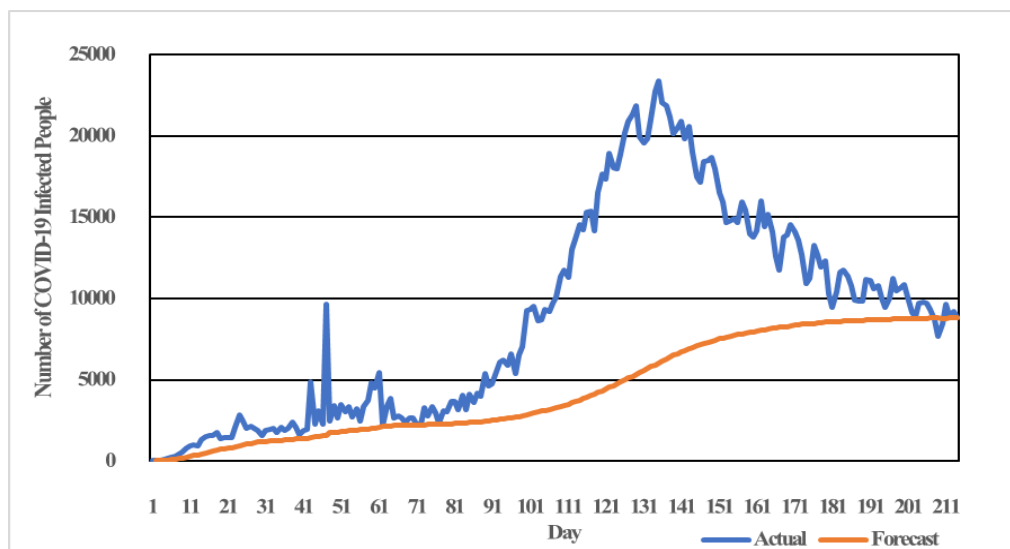


Figure 2 Line chart for COVID-19 Infected People in Thailand and forecast using global constant mean model

3.2.1 ตัวแบบค่าเฉลี่ยคงที่ระยะยาว

การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยด้วยตัวแบบค่าเฉลี่ยคงที่ระยะยาว แสดงผลการพยากรณ์ด้วยแผนภูมิเส้น แสดงดังภาพที่ 2

3.2.2 ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น

การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยด้วยตัวแบบค่าเฉลี่ยคงที่ระยะสั้น ใช้วิธีการพยากรณ์ 2 วิธี ได้แก่ ค่าเฉลี่ยเคลื่อนที่ชั้นเดียว (single moving average) และการทำให้เรียบแบบเอกซ์โพเนนเชียลชั้นเดียว (single exponential smoothing) แผนภูมิเส้นผลการพยากรณ์ แสดงดังภาพที่ 3 และ 4

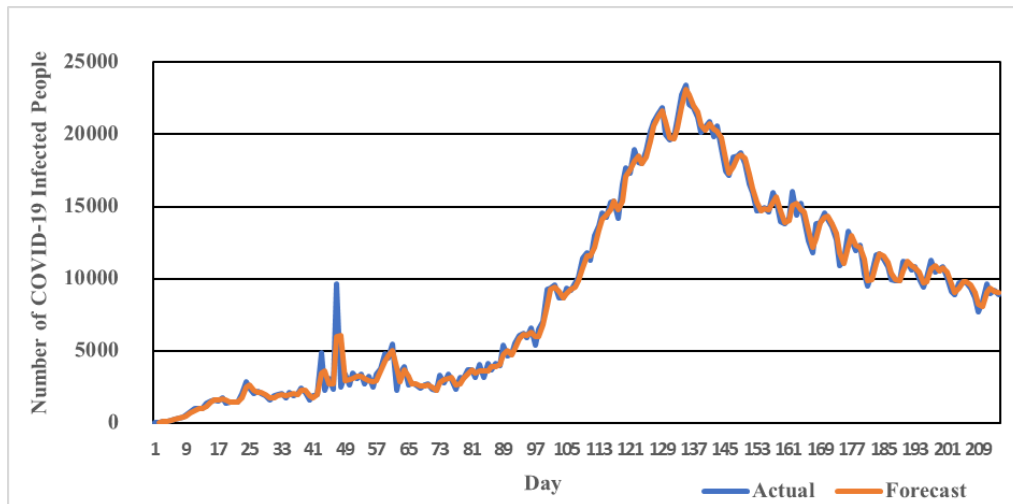


Figure 3 Line chart for COVID-19 Infected People in Thailand and forecast using local constant mean model: single moving average with $n = 2$

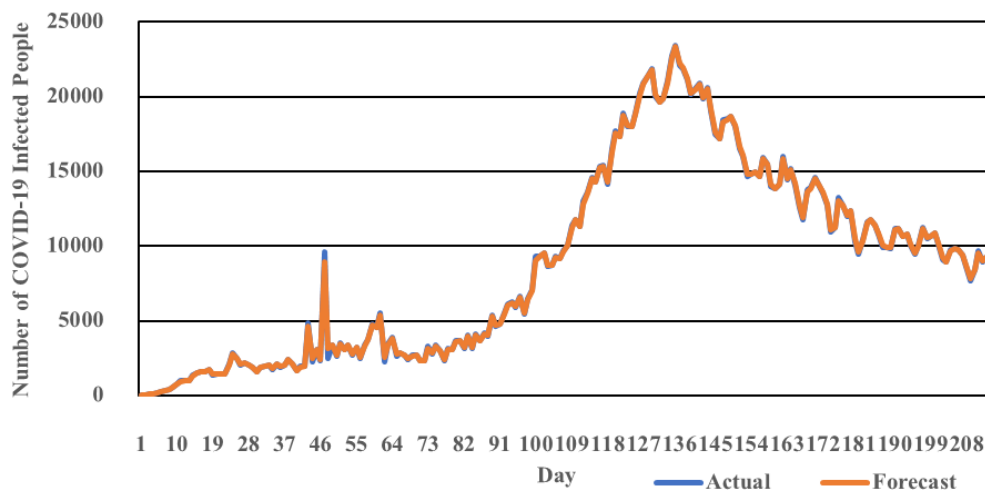


Figure 4 Line chart for COVID-19 Infected People in Thailand and forecast using local constant mean model: single exponential smoothing with α (smoothing constant) = 0.1

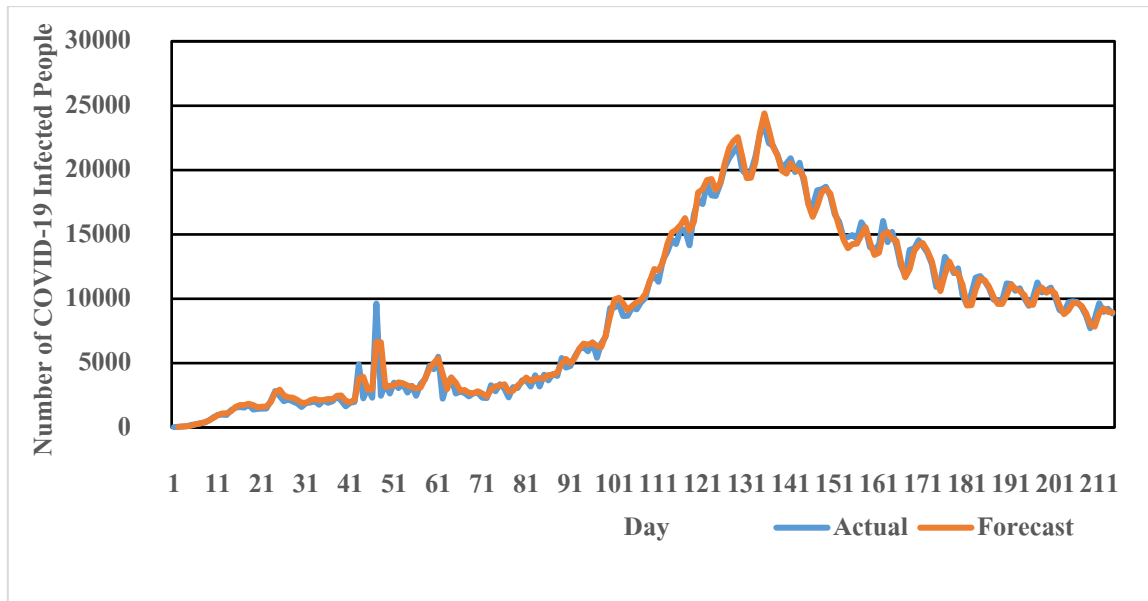


Figure 5 Line chart for COVID-19 Infected People in Thailand and forecast using adaptive forecasting model: adaptive filtering

3.2.3 การพยากรณ์แบบปรับได้

การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยด้วยเทคนิคการพยากรณ์แบบปรับได้ โดยวิธีการพยากรณ์แบบการกรองที่ปรับได้ แสดงดังภาพที่ 5

3.2.4 Box - Jenkins

การพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยด้วยตัวแบบการพยากรณ์ของ Box - Jenkins

พบว่า ตัวแบบพยากรณ์ที่เหมาะสมกับข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 คือ ตัวแบบรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่หรือตัวแบบอาร์มา (Autoregressive Integrated Moving Average Model: ARIMA model) ที่มีค่า $p = 5$, $q = 3$ และ $d = 2$ หรือ ตัวแบบ ARIMA(5, 2, 3) และจากการประมาณค่าพารามิเตอร์ต่าง ๆ ของตัวแบบดังกล่าว สามารถเขียนเป็นสมการพยากรณ์จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย ได้ดังนี้

$$Y_t = 0.5053Y_{t-1} - 0.5415Y_{t-2} - 0.4773Y_{t-3} - 0.0555Y_{t-4} - 0.3101Y_{t-5} + u_t - 1.9963u_{t-1} + 1.8733u_{t-2} - 0.7693u_{t-3}$$

เมื่อ Y_t คือ ผลต่างที่สองของข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19

$$\text{โดยที่ } Y_t = X_t - 2X_{t-1} + X_{t-2}$$

ทั้งนี้ เกณฑ์สารสนเทศของอะกะอิกะหรือเอไอซี (Akaike's Information Criterion: AIC) ของตัวแบบ ARIMA (5, 2, 3) พบว่า ค่า AIC มีค่าเท่ากับ 3514.41

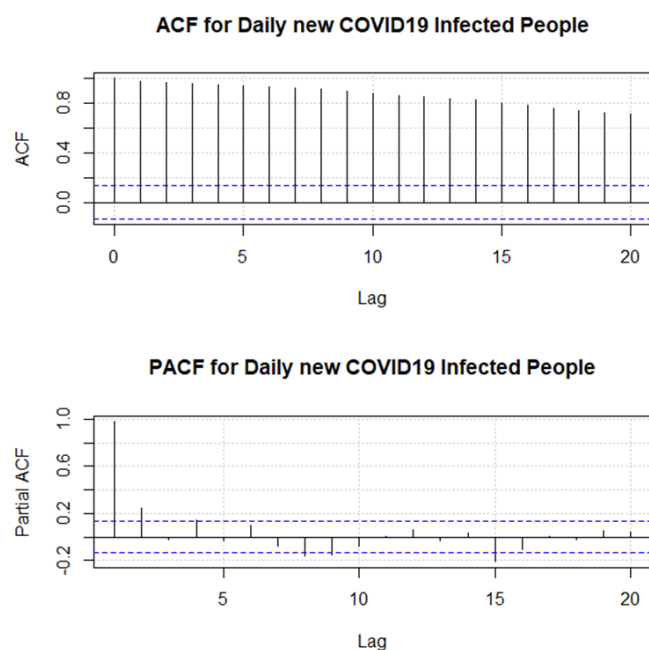
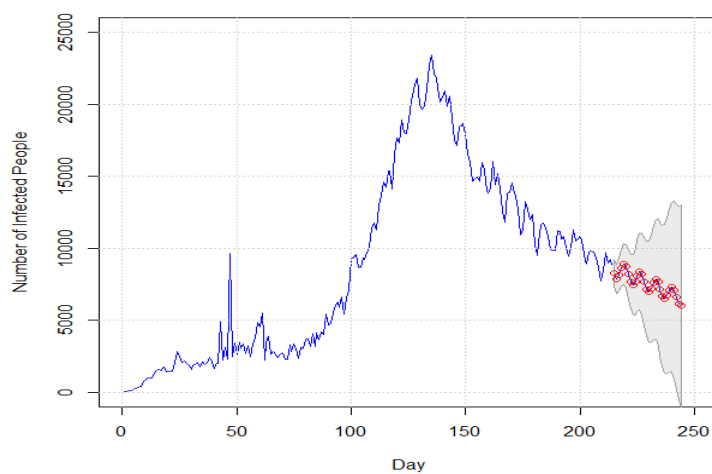
3.3 ความคลาดเคลื่อนที่เกิดจากการพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย

ความคลาดเคลื่อนที่เกิดจากการพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยด้วยเทคนิคการพยากรณ์ต่าง ๆ แสดงได้ดังตารางที่ 2 โดยการ

พยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย ที่มีค่า MAD, MSE และ MAPE น้อยที่สุด 3 อันดับแรก ได้แก่ การทำให้เรียบแบบเอกซ์โพเนนเชียลชั้นเดียว เมื่อค่าคงที่การทำให้เรียบ (α) เท่ากับ 0.1, 0.5 และวิธีค่าเฉลี่ยเคลื่อนที่ชั้นเดียว เมื่อจำนวนเทอม (n) เท่ากับ 2 (ช่องสีเหลือง สีส้มและสีน้ำเงิน ตามลำดับ) และการพยากรณ์ด้วยตัวแบบค่าเฉลี่ยคงที่ระยะยาวมีค่า MAD, MSE และ MAPE มากที่สุด (ช่องสีเขียว)

Table 2 MAD, MSE and MAPE of statistical forecasting of COVID-19 infected people in Thailand

Forecasting Technique	MAD	MSE	MAPE
global constant mean model	4,386.42	42,020,383	43.95
local constant mean model			
1) single moving average when $n = 2$	354.05	308,654.08	6.76
2) single exponential smoothing when $\alpha = 0.1$	69.52	11,638.21	1.30
single exponential smoothing when $\alpha = 0.5$	361.67	283,083.80	6.37
adaptive forecasting model: adaptive filtering	388.58	344,299.75	7.28
ARIMA (5, 2, 3)	727.10	682,359.25	11.73

**Figure 6** Autocorrelation function: ACF and partial autocorrelation function: PACF for COVID-19 Infected People in Thailand**Figure 7** Line chart for COVID-19 Infected People in Thailand and forecast using ARIMA (5, 2, 3) between November 1 - 30, 2021

4. สรุป

การวิจัยนี้ได้นำเสนอผลการพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย รวมถึงเปรียบเทียบค่าพยากรณ์และค่าที่แท้จริง พร้อมทั้งนำเสนอความคลาดเคลื่อนที่เกิดจากการพยากรณ์ทางสถิติด้วยตัวแบบต่าง ๆ ข้อมูลที่นำมาใช้ในการวิจัย คือ ข้อมูลทุติยภูมิ ซึ่งเป็นข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 จากการรายงานของศูนย์บริหารสถานการณ์โควิด-19 (ศบค.) ระหว่างวันที่ 1 เมษายน 2564 - 31 ตุลาคม 2564 เป็นเวลา 7 เดือน หรือ 214 วัน โดยข้อมูลที่ใช้ในการพยากรณ์ครั้งนี้จัดเป็นข้อมูลอนุกรมเวลาที่มีการจัดเก็บอย่างต่อเนื่องทุกวัน

จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยมีค่าต่ำสุดในเดือนเมษายน เท่ากับ 26 คน และมีค่าสูงสุดในเดือนสิงหาคม เท่ากับ 23,418 คน และเมื่อพิจารณาจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเฉลี่ยพบว่ามีค่าต่ำสุดในเดือนเมษายน เท่ากับ 1,209 คน และมีค่าสูงสุดในเดือนสิงหาคม เท่ากับ 19,595 คน ทั้งนี้จำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเฉลี่ย 7 เดือน เท่ากับ 8,802 คน

การพยากรณ์ทางสถิติจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย ที่ใช้ในการศึกษา ได้แก่ ตัวแบบค่าเฉลี่ยคงที่ระยะยาว ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น ด้วยวิธีค่าเฉลี่ยเคลื่อนที่ชั้นเดียว เมื่อจำนวนเทอม (n) เท่ากับ 2 และการทำให้เรียบแบบเอกซ์โพเนนเชียลชั้นเดียว เมื่อค่าคงที่การทำให้เรียบ (α) เท่ากับ 0.1 และ 0.5 การพยากรณ์แบบปรับได้ โดยวิธีการพยากรณ์แบบการกรองที่ปรับได้ และ Box - Jenkins โดยตัวแบบที่เหมาะสมสำหรับการพยากรณ์ คือ ตัวแบบรวม การถดถอยในตัวกับการเฉลี่ยเคลื่อนที่หรือตัวแบบอาร์มา (Autoregressive Integrated Moving Average Model: ARIMA model) ที่มีค่า $p = 5$, $q = 3$ และ $d = 2$ หรือตัวแบบ ARIMA (5, 2, 3)

การศึกษาความคลาดเคลื่อนที่เกิดจากการพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย โดย คำนวณค่าความคลาดเคลื่อนจากการพยากรณ์ ได้แก่ ค่าเบี่ยงเบนสัมบูรณ์เฉลี่ย (Mean Absolute Deviation, MAD) ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error, MSE) และร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ค่าต่ำที่สุด (Mean Absolute Percentage Error, MAPE) พบว่า การพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทย ที่มีค่า MAD, MSE และ MAPE น้อยที่สุด 3 อันดับแรก ได้แก่ การทำให้เรียบแบบเอกซ์โพเนนเชียลชั้นเดียว เมื่อค่าคงที่การทำให้เรียบ (α) เท่ากับ 0.1, 0.5 และวิธีค่าเฉลี่ยเคลื่อนที่ชั้นเดียว เมื่อจำนวนเทอม (n) เท่ากับ 2 และการพยากรณ์

ด้วยตัวแบบค่าเฉลี่ยคงที่ระยะยาวมีค่า MAD, MSE และ MAPE มากที่สุด

เมื่อพิจารณาผลการศึกษานางานวิจัยนี้กับงานวิจัยที่เกี่ยวข้อง พบว่า วิธีการศึกษาและผลสรุปมีความแตกต่างกัน ซึ่งงานวิจัยที่เกี่ยวข้องเป็นการวิจัยที่ใช้วิธีการหรือตัวแบบทางคณิตศาสตร์ หรือตัวแบบอื่น ๆ ทางระบาดวิทยาในการศึกษา โดยใช้ข้อมูลจำนวนผู้ติดเชื้อที่หลากหลายหลายมิติ เช่น จำนวนผู้ติดเชื้อสะสม จำนวนผู้เสียชีวิตและรักษาหายในช่วงเวลาและข้อจำกัดที่แตกต่างกัน ในบางงานวิจัยมีการศึกษาข้อมูลผู้ป่วยจากประเทศอื่น ๆ เพื่อเปรียบเทียบกับประเทศไทย สำหรับการพยากรณ์ทางสถิติของจำนวนผู้ติดเชื้อไวรัสโควิด-19 ในประเทศไทยเรื่องนี้ ได้นำข้อมูลจำนวนผู้ติดเชื้อไวรัสโควิด-19 รายวัน ซึ่งถูกนำเสนอเป็นประจำวันโดยศูนย์บริหารสถานการณ์โควิด-19 (ศบค.) มาใช้ในการศึกษาด้วยวิธีการพยากรณ์ทางสถิติ 4 เทคนิค ได้แก่ ตัวแบบค่าเฉลี่ยคงที่ระยะยาว ตัวแบบค่าเฉลี่ยคงที่ระยะสั้น การพยากรณ์แบบปรับได้ และ Box - Jenkins และทำการศึกษาความคลาดเคลื่อนที่เกิดจากการพยากรณ์โดยใช้เกณฑ์ MAD MSE และ MAPE ซึ่งได้ผลการศึกษาก็แตกต่างกันอย่างสิ้นเชิง

5. เอกสารอ้างอิง

- Areepong, Y. and Sunthornwat, R. 2021. Forecasting modeling of the number of cumulative COVID-19 cased with deaths and recoveries removal in Thailand. **Science, Engineering and Health Studies** 15: 21020004.
- Ausawanlop, N. and Aeiutrakul, K. 2021. **The Mathematical Model to forecast of COVID-19 Pandemic.** Available Source: <https://thaipublica.org/2021/05/epidemic-sir-model/>, November 28, 2021. (in Thai)
- Faculty of Medicine Siriraj Hospital. 2021. **What is COVID-19?** Available Source: <https://www.gj.mahidol.ac.th/main/covid19/covid19is/>, October 25, 2021. (in Thai)
- Ingpansatawong, S. 2012. **Statistical Forecasting Techniques.** Khon Kaen University Press, Khon Kaen. (in Thai)
- Lorjirashunkul, W. and Jittaweij, J. 2005. **Forecasting Techniques.** National Institute of Development Administration, Bangkok. (in Thai)

- Ministry of Public Health. 2021. **Coronavirus disease 2019 (COVID-19)**. Available Source: https://www.who.int/docs/default-source/searo/thailand/2020-03-26-tha-sitrep-33-covid19-th-final.pdf?sfvrsn=8f5738e_0, October 25, 2021. (in Thai)
- Nuanchui, T. 2021. COVID-19 cumulative case forecasting by Deep Learning. Master Thesis of Science, Chulalongkorn University.
- Panichayakul, T. 2020. **Forecasting of COVID-19 Infected People Using Principles of Statistics**. Available Source: <https://www.sci.psu.ac.th/news/2020/04/coronavirus-prophecy-via-statistical-principles/>, October 25, 2021. (in Thai)
- Sinsomboonthong, S. 2016. **Data Analysis using MINITAB for Windows**. Chamchuri Product, Bangkok. (in Thai)
- Vorathamthongdee, S. and Congstitvatana, P. 2023. Predictive Analysis of COVID-19 Patients in Thailand using Multiple Countries Data. **ECTI Transaction on Application Research and Development** 3(1): 248647.
- Wongsathan, R. 2021. Real-time Prediction of the COVID-19 Epidemic in Thailand Using Simple Model – Free Method and Time Series Regression Model. **Walailak Journal of Science and Technology** 18(14): 10028.