วารสารวิชาการสาธารณสุข
ปีที่ 35 ฉบับที่ 2 มีนาคม - เมษายน 2569

Journal of Health Science of Thailand
Vol. 35 No. 2, March - April 2026

นิพนธ์ต้นฉบับ *Original article*

การพยากรณ์จำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอด: กรณีศึกษาโรงพยาบาลในจังหวัดเชียงใหม่

ไพรชา สุทนต์ วท.ม.*
ฐาปกรณ์ ไตรยะวิภาค วท.ม.*
วัฒนา ชยธวัช ปร.ด.**
** คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่ จังหวัดเชียงใหม่*
*** คณะสหเวชศาสตร์ มหาวิทยาลัยปทุมธานี จังหวัดปทุมธานี*
ติดต่อผู้เขียน: ไพรชา สุทนต์ Email: priracha_sut@g.cmru.ac.th

วันรับ: 23 ก.พ. 2568
วันแก้ไข: 6 พ.ย. 2568
วันตอบรับ: 20 พ.ย. 2568

บทคัดย่อ โรงพยาบาลในจังหวัดเชียงใหม่แห่งหนึ่งมีสถิติจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอดรายเดือนตั้งแต่ปีงบประมาณ 2558 จนถึงเดือนกันยายน 2566 เดือนละมากกว่า 4,000 คน ซึ่งเป็นข้อมูลอนุกรมเวลาที่มีส่วนประกอบของฤดูกาลและแนวโน้มเพิ่มขึ้น การวิจัยนี้มีวัตถุประสงค์เพื่อทำการพยากรณ์ด้วยการแยกตัวประกอบแบบดั้งเดิมแบบจำลองการคูณที่คำนวณดัชนีฤดูกาลด้วยวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ (classical decomposition ratio to moving average: CDRMA) และวิธีบ๊อกซ์และเจนกินส์แบบจำลอง seasonal autoregressive integrated moving average (SARIMA) กับการพยากรณ์ผสมทฤษฎีแบบจำลองเกรย์กับดัชนีฤดูกาล โดยใช้แบบจำลองเกรย์พยากรณ์จำนวนผู้ป่วยในรายปีที่มีค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (mean absolute percentage error: MAPE) น้อยที่สุดแล้วแจกแจงค่าพยากรณ์รายเดือนด้วยดัชนีฤดูกาลวิธีค่าเฉลี่ยอย่างง่าย (GSA) กับวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ (GRMA) โดยใช้ข้อมูลจำนวนผู้ป่วยในรายเดือนที่รวบรวมจากระบบรายงานข้อมูล CMI ทีมพัฒนาโปรแกรมภายใต้สังกัดสำนักงานสาธารณสุขจังหวัดเชียงใหม่ ระหว่างปีงบประมาณ 2557 - 2566 ผลการศึกษาพบว่า แบบจำลองที่เหมาะสม คือ ARIMA$(1,1,1)(0,1,1)_{12}$ ซึ่งมีค่า MAPE จากอดีต 96 เดือน ร้อยละ 4.44 ต่ำกว่าแบบจำลอง CDRMA ที่มี MAPE ร้อยละ 5.39 ส่วนการพยากรณ์แบบผสมตัวแบบเกรย์จากการพยากรณ์รายปีที่เหมาะสมคือ GM(1,1) Error Periodic Correction มีค่า MAPE ร้อยละ 1.18 เมื่อนำมาแจกแจงเป็นค่าพยากรณ์รายเดือนของปี 2566 แล้วเปรียบเทียบค่า MAPE 9 เดือนที่เทียบกับข้อมูลจริง พบว่าแบบจำลอง GRMA และ GSA มีค่า MAPE ร้อยละ 4.88 และ 4.05 ตามลำดับ เนื่องจากข้อมูลมีส่วนประกอบฤดูกาลและมีแนวโน้มเพิ่มขึ้น จึงเลือกแบบจำลอง GRMA การตรวจสอบความแม่นยำของการพยากรณ์ได้จากรายงานสถิติสาธารณสุขของจังหวัดเชียงใหม่ รายงานที่ 178 ต้นปี 2567 พบว่ามีจำนวน 56,937 ราย ซึ่งวิธี GRMA ที่พยากรณ์ด้วย GM(1,1)EPC สูงกว่าค่าจริงร้อยละ 2.05

คำสำคัญ: **การพยากรณ์; ผู้ป่วยใน; โรคฝุ่นจับปอด**

บทนำ

กลุ่มโรคฝุ่นจับปอด (pneumoconiosis) เป็นกลุ่มของโรคปอด ซึ่งสัมพันธ์กับของฝุ่นที่สูดดมเข้าไปและปฏิกิริยาของเนื้อเยื่อปอด ซึ่งนำไปสู่การบาดเจ็บที่ปอดอย่างถาวร เนื่องจากขาดการป้องกันฝุ่นในที่ทำงาน ความล้มเหลวในการวินิจฉัยโรคตั้งแต่เนิ่น ๆ และการรักษาโรคที่มีประสิทธิผลอย่างจำกัด โรคฝุ่นจับปอดจึงยังคงเป็นปัญหาสาธารณสุขระดับโลกที่ร้ายแรง[1]

การจำแนกกลุ่มโรคฝุ่นจับปอดในระบบการจำแนกโรคระหว่างประเทศ ฉบับที่ 10 (International Classification of Diseases, 10th Revision: ICD-10)[2] อยู่ในรายการรหัสโรค JXX ดังนี้ J60 ฝุ่นจับปอดในคนงานเหมืองถ่านหิน J61 แอสเบสโตสิส (โรคปอดใยหิน) แร่ใยหิน J62 โรคปอดฝุ่นหิน แป้งฝุ่น (talc) ฝุ่นหิน (silica) J63 ฝุ่นจับปอดจากฝุ่นอนินทรีย์อื่น ๆ อลูมิเนียม บอกไซต์ เบริลเลียม กราไฟต์ ฝุ่นเหล็ก ฝุ่นและฟูมของดีบุก และฝุ่นผสม J65 สำหรับอาการป่วยด้วยโรคฝุ่นจับปอดร่วมกับวัณโรคปอด ไม่ว่าจะเป็นโรคฝุ่นจับปอดชนิดใดก็ตาม J84 โลหะหนัก (โคบอลต์) J90 J92 J94 เหมือนกับ J61 ตามสถิติการป่วยด้วยกลุ่มโรคฝุ่นสะสมในปอดของจังหวัดเชียงใหม่ ระหว่างปีงบประมาณ 2557-2566 ทีมพัฒนาสารสนเทศ เขตบริการสุขภาพที่ 1[3] พบว่า รายงานจำนวนผู้ป่วยในแตกต่างไปจากรายงานกระทรวงสาธารณสุขในระบบคลังข้อมูลด้านการแพทย์และสุขภาพ (Health Data Center: HDC) อัตราป่วยโรคระบบทางเดินหายใจ เขตสุขภาพที่ 1 จังหวัดเชียงใหม่ โรงพยาบาลนครพิงค์[4] ที่รวบรวมจากจำนวนประชากรในเขตรับผิดชอบป่วยโรคระบบทางเดินหายใจ (ICD-10 TM CODE J00-J39, J60-J99) ซึ่งอาจจะเกิดจากการรายงานเข้าสู่ระบบของหน่วยงานในเวลาที่แตกต่างกัน จึงเลือกใช้รายงานของเขตบริการสุขภาพที่ 1 ที่กำกับการอยู่ในพื้นที่

จำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอดรายใหม่ของประเทศจีนเพิ่มขึ้นมากกว่า 20,000 คนต่อปี มีการพยากรณ์จำนวนผู้ป่วยรายใหม่เพื่อเป็นข้อมูลในการวางแผนลดจำนวนผู้ป่วยรายใหม่ด้วยการพยากรณ์อนุกรมเวลาแบบจำลอง autoregressive integrated moving average (ARIMA) แบบจำลอง deep neural networks (DNN) และแบบจำลอง long short-term memory neural network (LSTM) หลายตัวแปร ตัวชี้วัดการประเมินประสิทธิภาพความแม่นยำ ได้แก่ รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (root mean squared error: RMSE) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error: MAE) และค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดสัมบูรณ์ (mean absolute percentage error: MAPE) ใช้เปรียบเทียบผลการทำนายของทั้ง 3 โมเดล[5] วิธีการพยากรณ์อนุกรมเวลาเป็นวิธีการคาดการณ์และการพยากรณ์เพื่อสะท้อนแนวโน้มการพัฒนาของสิ่งต่าง ๆ ผ่านอนุกรมเวลาที่เป็นข้อมูลในอดีต[6] วิธีการทำนายอนุกรมเวลาแบบดั้งเดิมทั่วไป ได้แก่ แบบจำลองค่าเฉลี่ยเคลื่อนที่แบบรวมอัตโนมัติแบบถดถอย (ARIMA) และวิธีการปรับให้เรียบแบบเอกซ์โปเนนเชียลของ Holt-Winters ซึ่งแบบจำลอง ARIMA นั้นเป็นแบบจำลองคลาสสิกและได้รับความนิยมมากที่สุด[7,8] แบบจำลอง ARIMA เกี่ยวข้องกับค่าคงที่ของการเปลี่ยนแปลงแนวโน้ม การรบกวนแบบสุ่ม การเปลี่ยนแปลงตามระยะเวลา และตัวแปรสุ่มอื่น ๆ ที่เกี่ยวข้องในกระบวนการวิเคราะห์อนุกรมเวลา เนื่องจากข้อดีของโครงสร้างที่เรียบง่าย การนำไปใช้งานที่แข็งแกร่ง และความสามารถในการตีความชุดข้อมูล แบบจำลอง ARIMA จึงถูกนำไปใช้อย่างประสบความสำเร็จในสาขาการแพทย์และสุขภาพที่ผ่านมา[9] นอกจากนี้ยังมีการพยากรณ์อัตราการติดเชื้อในระบบทางเดินหายใจประจำปีของโรงพยาบาลระดับ 3 ในเมือง Yan'an ประเทศจีน โดยใช้ข้อมูลที่รวบรวมตั้งแต่ พ.ศ. 2554 - 2560 แบบจำลอง GM(1,1) ใช้ในการทำนายและใช้เปอร์เซ็นต์ความคลาดเคลื่อนเฉลี่ยสัมบูรณ์เพื่อประเมินความแม่นยำในการทำนายของแบบจำลอง ซึ่งผลการวิจัยพบว่า แบบจำลองการทำนายทางสถิติ GM(1,1) ได้รับการสร้างขึ้นโดยมีระดับความเหมาะสมที่ดีและความน่าเชื่อถือสูงของการทำนายการคาด-

การณ์[10] ใช้ข้อมูล พ.ศ. 2543 – 2562 พยากรณ์ผู้ป่วย พ.ศ. 2563 – 2564 เนื่องจากข้อมูลมีลักษณะโค้งขึ้นลงเป็นรูปตัว S ที่แบบจำลอง GM(1,1) ใช้พยากรณ์ได้ด้อยกว่า Generalized Additive Mode (GAM)[11] อย่างไรตาม ทฤษฎีแบบจำลองระบบเกรย์ก็ยังมีตัวแบบอื่นที่ปรับค่าให้สอดคล้องกับรูปแบบของข้อมูลด้วย[12] การพยากรณ์ไม่ใช่แค่การคาดการณ์ตัวเลข แต่เป็นเครื่องมือที่ช่วยให้ผู้บริหารโรงพยาบาล “มองเห็นอนาคต” และ “เตรียมพร้อม” เพื่อให้การดูแลผู้ป่วยเป็นไปอย่างมีประสิทธิภาพและปลอดภัย

โรงพยาบาลนครพิงค์เป็นโรงพยาบาลศูนย์ประจำจังหวัด จังหวัดเชียงใหม่ สังกัดกระทรวงสาธารณสุข[13] มีข้อมูลจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอดสูงถึง 56,000 คน ในปี พ.ศ. 2565 จึงมีวัตถุประสงค์ศึกษาเปรียบเทียบการพยากรณ์จำนวนผู้ป่วยในรายเดือนสำหรับ พ.ศ. 2566 ด้วยวิธีการแยกตัวประกอบ (decomposition) การพยากรณ์ผสมทฤษฎีแบบจำลองเกรย์กับดัชนีฤดูกาล และแบบจำลอง SARIMA (seasonal autoregressive integrated moving average) เนื่องจากเป็นแบบจำลองที่สามารถคำนวณด้วยโปรแกรม MICROSOFT Excel[14,15] ที่มีอยู่กับเครื่องคอมพิวเตอร์ และโปรแกรม R กับ Package Forecast โปรแกรมฟรีที่สามารถดาวน์โหลดมาใช้งานได้ไม่ยุ่งยากนัก[16,17] เพื่อใช้เป็นแนวทางการพยากรณ์เพื่อการวางแผนบริการผู้ป่วยในที่เหมาะสมต่อไป

วิธีการศึกษา

การศึกษานี้เป็นการวิจัยเชิงปริมาณโดยใช้แบบจำลองการพยากรณ์ข้อมูลอนุกรมเวลา (quantitative research using time series forecasting models)

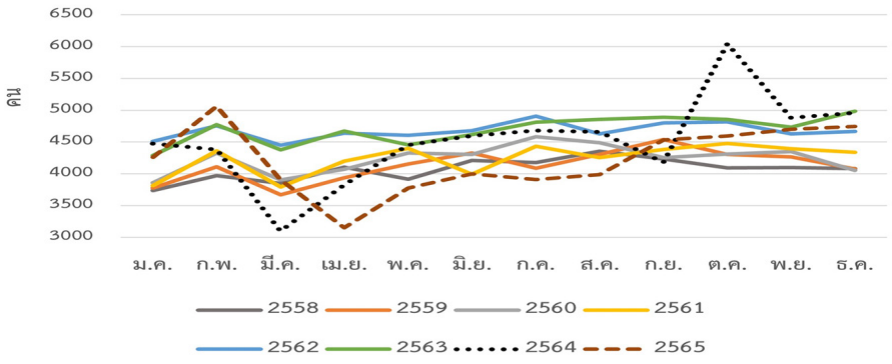
การรวบรวมข้อมูล

ข้อมูลที่นำมาศึกษาเป็นจำนวนผู้ป่วยในโรคฝุ่นจับปอดรายเดือนรวบรวมจากระบบรายงานข้อมูล CMI ทีมพัฒนาโปรแกรมภายใต้สังกัดสำนักงานสาธารณสุขจังหวัดเชียงใหม่ วันที่ประมวลผล 20 ตุลาคม พ.ศ. 2566[3] รายงานที่ 178 ร้อยละการป่วยด้วยกลุ่มโรคฝุ่นสะสมในปอดของโรงพยาบาลนครพิงค์ จังหวัดเชียงใหม่ จำนวนผู้ป่วยในที่จำหน่ายทั้งหมดรายเดือนระหว่างปีงบประมาณ 2557 – 2566 ได้แปลงเป็นข้อมูลตามปีปฏิทินระหว่าง พ.ศ. 2558 – 2565 โดยในปี พ.ศ. 2559 พบผู้ป่วยโรคนี้น้อยที่สุด 49,088 คน มากที่สุด 56,742 คนในปี พ.ศ. 2563 และเฉลี่ยปีละ 52,745 คน

การพิจารณารูปแบบข้อมูล

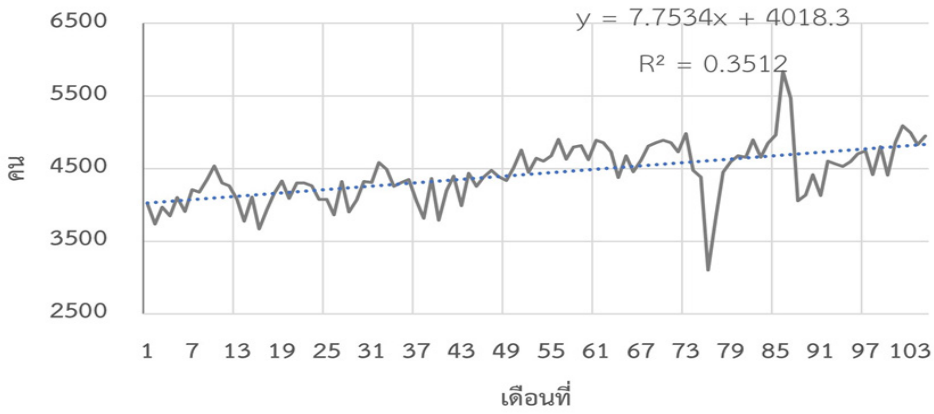
จำนวนผู้ป่วยในโรคฝุ่นจับปอดรายเดือนของโรงพยาบาลเป็นข้อมูลรายเดือนจำนวน 8 ปี[3] เมื่อนำมาพล็อตกราฟเส้นรายเดือน 12 เดือน ซ้อนกัน 8 ปี พบว่ามีบางปีมีจำนวนผู้ป่วยในรายเดือนคงที่ หลายปีมีรูปแบบซ้ำกันในรอบ 12 เดือนเป็นรูปแบบของฤดูกาลกล่าวคือลดลงในเดือนมีนาคม เพิ่มสูงขึ้นเล็กน้อยประมาณเดือนกันยายนดังแสดงในภาพที่ 1 และเมื่อนำมาพล็อต

ภาพที่ 1 รูปแบบข้อมูลจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นสะสมในปอดของโรงพยาบาลนครพิงค์ จังหวัดเชียงใหม่รายเดือน พ.ศ. 2558 – 2565



ต่อเนื่องกัน 105 เดือน พบว่า มีแนวโน้มเพิ่มขึ้น (ตามสมการเส้นตรง y = 7.7534x+4018.3) มีค่าความชัน 7.7534 หมายความว่า เวลาเพิ่มขึ้น 1 เดือนมีผู้ป่วยเพิ่มขึ้น 7.7534 คน ดังแสดงในภาพที่ 2

ภาพที่ 2 รูปแบบแนวโน้มข้อมูลจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นสะสมในปอดของโรงพยาบาลนครพิงค์ จังหวัดเชียงใหม่รายเดือน (105 เดือน พ.ศ. 2558 ถึง ก.ย. พ.ศ. 2566)



การเลือกวิธีการพยากรณ์ที่สอดคล้องกับรูปแบบข้อมูล

จากรูปแบบข้อมูลมีลักษณะของฤดูกาลรูปแบบซ้ำ ๆ กันทุกปี ยกเว้น พ.ศ. 2564 และ 2565 ที่มีจำนวนผู้ป่วยต่ำในเดือนมีนาคมและเมษายน ตามลำดับ ขณะที่ปี พ.ศ. 2564 มีผู้ป่วยสูงมากในเดือนตุลาคม และข้อมูลรายเดือนก็มีแนวโน้มเพิ่มขึ้นด้วย จึงเลือกการพยากรณ์จำนวนผู้ป่วยรายเดือน พ.ศ. 2566 ดังนี้

การพยากรณ์ด้วยวิธีการแยกส่วนประกอบแบบดั้งเดิม

แบบจำลองการคูณคำนวณดัชนีฤดูกาลด้วยวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ (ratio to moving average) ซึ่งเป็นวิธีที่สอดคล้องกับรูปแบบข้อมูลที่มีองค์ประกอบของฤดูกาลและแนวโน้ม[18]

การพยากรณ์ด้วยทฤษฎีระบบเกรย์[12]

ทฤษฎีระบบเกรย์ (Grey System Theory) เหมาะสมกับข้อมูลที่มีจำกัด กรณีจำนวนผู้ป่วยในโรคฝุ่นจับปอดนี้มีข้อมูลครบปีเพียง 7 ปี เนื่องจากทฤษฎีนี้ถูกออกแบบมาเพื่อจัดการกับข้อมูลที่ไม่สมบูรณ์และ/หรือมีจำนวนน้อย ที่ไม่สามารถนำไปใช้กับวิธีการทางสถิติปกติได้ โดยข้อได้เปรียบและเหตุผลที่ทฤษฎีระบบเกรย์เหมาะสมมีดังนี้[12]:

1) สามารถวิเคราะห์แนวโน้มในข้อมูลชุดเล็กได้ ทฤษฎีระบบเกรย์สามารถสร้างแบบจำลองที่แม่นยำ แม้ว่าจะมีข้อมูลที่จำกัดเพียง 4-10 จุดเท่านั้น

2) ลดความซับซ้อนของข้อมูล โดยข้อมูลที่อาจดูไม่ชัดเจนหรือลำดับสุ่มจะถูกเปลี่ยนเป็นรูปแบบเชิงเส้นผ่านกระบวนการสะสม (accumulated generation operation: AGO) ทำให้ข้อมูลมีลักษณะเพิ่มขึ้นอย่างเดียว (monotonic increasing) เพื่อหาความสัมพันธ์และแนวโน้มได้ง่ายขึ้น

3) ทฤษฎีระบบเกรย์สามารถพยากรณ์แนวโน้มข้อมูลในอนาคตในปีถัดจากชุดข้อมูลได้

4) ความยืดหยุ่นสูง ไม่ต้องการสมมติฐานที่เข้มงวด เช่น การแจกแจงปกติหรือความสัมพันธ์เชิงเส้นตรงในข้อมูล

การพยากรณ์แบบผสม

การพยากรณ์จำนวนผู้ป่วย พ.ศ. 2566 ด้วยแบบจำลองจากทฤษฎีระบบเกรย์ ซึ่งประกอบไปด้วยแบบจำลอง GM(1,1), GM(1,1) error periodic correction: GM(1,1) EPC และ Grey Verhulst Model (GVM) แล้วเลือกแบบจำลองที่มีความแม่นยำสูงแล้ว จึงนำค่า

พยากรณ์รายปีมาแจกแจงเป็นรายเดือนด้วยดัชนีฤดูกาล เนื่องจากทฤษฎีระบบเกรย์ใช้กับข้อมูลจำนวนน้อย ทฤษฎีระบบเกรย์ใช้ในการพยากรณ์อนุกรมเวลา และสามารถใช้กับข้อมูลจำนวนน้อยที่สุดเพียง 4 คาบเวลาก็สามารถใช้ทำนายคาบเวลาที่ 5 ได้ แบบจำลองเริ่มต้นคือ GM(1,1) (first order grey model for one variable) ซึ่งจะเริ่มต้นจากการแปลงข้อมูลจริงที่ศึกษาเป็นข้อมูลสะสม แล้วใช้สมการถดถอยเชิงเส้นตรงประมาณค่าคงที่และค่าความชัน จากนั้นทำนายค่าสะสมแล้วหักออกด้วยค่าสะสมที่ทำนายก่อนหน้าก็จะได้ค่าทำนายของแต่ละคาบเวลา ซึ่งกล่าวได้ว่าเหมาะสมกับข้อมูลอนุกรมเวลาที่มีจำนวนค่าสังเกตเพิ่มสูงขึ้นทุกปี นอกจากนี้ยังมีแบบจำลองที่มีการแก้ไขข้อผิดพลาดเป็นระยะเพื่อให้เหมาะกับข้อมูลที่มีรูปแบบเพิ่มขึ้นและลดลงเป็นโค้งรูปตัว S ด้วยการพิจารณาว่าแบบจำลองการทำนายสามารถทำนายได้แม่นยำเพียงใดก็สามารถพิจารณาได้จากค่าทำนายที่เบี่ยงเบนไปจากค่าจริง หรือค่าสหสัมพันธ์ของค่าจริงกับค่าทำนาย[12]

การคำนวณดัชนีฤดูกาล (seasonal index)

วิธีการแยกส่วนประกอบแบบดั้งเดิมค่าเฉลี่ยอย่างง่าย (simple average) เนื่องจากข้อมูล 105 เดือน มีค่าเฉลี่ย 4,429 คนต่อเดือน ส่วนเบี่ยงเบนมาตรฐาน 397 คน สัมประสิทธิ์ความแปรปรวน ร้อยละ 8.95 ซึ่งพิจารณาได้ว่าข้อมูลรายเดือนวิ่งรอบ ๆ ค่าเฉลี่ยจึงใช้วิธีค่าเฉลี่ยอย่างง่าย (simple moving average) และวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ (ratio to moving average) สำหรับข้อมูลที่มีความผันแปรตามฤดูกาล มีแนวโน้มและวัฏจักรร่วมด้วย[19,20]

การพยากรณ์ด้วยวิธีบอกซ์-เจนกินส์[21]

แบบจำลอง SARIMA (Seasonal AutoRegressive Integrated Moving Average) เป็นแบบจำลองที่ครอบคลุมวิธีบอกส์-เจนกินส์ เนื่องจากมีข้อมูลรายเดือนมากถึง 8 ปี จึงใช้ความสามารถของวิธีการ SARIMA ในการพยากรณ์รายเดือนล่วงหน้า 12 เดือนโดยตรง แบบจำลอง SARIMA(p,d,q)(P,D,Q)m ซึ่งหมายถึงแบบจำลอง ARIMA (p, d, q) มีส่วนประกอบที่สำคัญ 3 ส่วน ได้แก่ auto regressive, AR (p), intergrated (I) และ moving average, MA (q) สำหรับ AR (p) เป็นรูปแบบที่แสดงว่าค่าสังเกต yt ขึ้นอยู่กับค่าของ yt-1,yt-2,..., yt-p หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) เป็นรูปแบบที่แสดงค่าสังเกต yt ขึ้นอยู่กับค่าความคลาดเคลื่อน et-1,et-2,...,et-q หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า ส่วน Integrated (I) เป็นการหาผลต่าง (difference) ของอนุกรมเวลา เหตุผลสำคัญที่ต้องหาผลต่างของอนุกรมเวลา เนื่องจากแบบจำลอง ARIMA ใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลาที่มีคุณสมบัติคงที่ (stationary) เท่านั้น ในกรณีที่ข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์สมบัติไม่คงที่ (non-stationary) จะต้องทำการแปลงข้อมูลอนุกรมเวลาดังกล่าวให้มีคุณสมบัติคงที่ก่อน โดยการหาผลต่างของข้อมูลอนุกรมเวลา SARIMA มีพารามิเตอร์ตามฤดูกาลอีกสามรายการ (P, D, Q) พารามิเตอร์เพิ่มเติมอีกสามรายการคำนึงถึงองค์ประกอบการถดถอยอัตโนมัติ (P) องค์ประกอบผลต่าง (D) และองค์ประกอบค่าเฉลี่ยเคลื่อนที่ (Q) ในระดับฤดูกาล โดยที่ m คือ จำนวนการค่าสังเกตในแต่ละฤดูกาล (ในกรณีนี้ m คือ 12) ส่วนประกอบตามฤดูกาลของโมเดลจะแสดงด้วยตัวพิมพ์ใหญ่ และส่วนประกอบที่ไม่ใช่ตามฤดูกาลของโมเดลจะแสดงด้วยตัวพิมพ์เล็ก[22] การหาค่าพารามิเตอร์สำหรับแบบจำลอง SARIMA โดยใช้ โปรแกรม R forecast package[23]

การพยากรณ์เปรียบเทียบ

การพยากรณ์จากจำนวนผู้ป่วยรายเดือน 2 วิธี คือ การแยกส่วนประกอบแบบดั้งเดิมวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ (classical decomposition ratio to moving average)[24] เรียกโดยย่อว่า CDRMA และ SARIMA (seasonal autoregressive integrated moving average)[23]

การพยากรณ์แบบผสมพยากรณ์มูลค่าการใช้ยารายปีด้วยแบบจำลองจากทฤษฎีระบบเกรย์ที่มีค่าความแม่นยำ

สูงสุด (MAPE ต่ำสุด) นำมาแจกแจงรายเดือนด้วยดัชนีฤดูกาลแบบค่าเฉลี่ยอย่างง่าย[25] เป็นการพยากรณ์แบบผสมทฤษฎีระบบเกรย์กับดัชนีฤดูกาลแบบค่าเฉลี่ยอย่างง่าย (hybrid forecasting using Grey System Theory and simple moving average seasonal index: GSA) และแบบผสมกับวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่[18] (hybrid forecasting using Grey System Theory and ratio to moving average seasonal index: GRMA)

เกณฑ์ตัดสินความแม่นยำของแบบจำลอง

ค่าข้อมูลจริง y_i จำนวน n ค่า โดย i=1, 2, ..., n ส่วน y_i เป็นค่าพยากรณ์ สำหรับคาบเวลาตามข้อมูลที่นำมาใช้พยากรณ์ i=1,2,...,n ส่วนค่าพยากรณ์ไปในอนาคต i=n+1,n+2,...

ค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดสัมบูรณ์ (The Mean Absolute Percentage Error - MAPE)[26]

$$MAPE = \left(\frac{1}{n} \sum_{k=1}^{n} \left| \frac{y_k - \hat{y}_k}{y_k} \right| \right) \times 100\%$$

MAPE หน่วยเป็นร้อยละ โดยมีเกณฑ์ความแม่นยำตามค่า MAPE[27] ดังนี้

ค่าน้อยกว่าร้อยละ 10 มีความแม่นยำสูง
ค่าร้อยละ 10 - 20 ใช้พยากรณ์ได้ดี
ค่าร้อยละ 20 - 50 มีเหตุผลพอที่จะใช้พยากรณ์
ค่ามากกว่าร้อยละ 50 ไม่มีความแม่นยำ

ผลการศึกษา

ดัชนีฤดูกาล

ดัชนีฤดูกาลโดยวิธีค่าเฉลี่ยอย่างง่าย วิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ คำนวณจากข้อมูลมูลค่าการใช้รายเดือนระหว่าง พ.ศ. 2558 - 2565 ดังแสดงในตารางที่ 1 ซึ่งค่าดัชนีฤดูกาลในเดือนใดเท่ากับ 1 หมายความว่าในเดือนนั้นมีจำนวนผู้ป่วยในมีจำนวนตามค่าเฉลี่ยรายเดือนตามปกติ กรณีที่มากกว่า 1 หมายความว่า ในเดือนนั้นมีจำนวนผู้ป่วยในมากกว่าค่าเฉลี่ยรายเดือน เช่น 1.25 = สูงกว่าค่าเฉลี่ยร้อยละ 25 กรณีที่น้อยกว่า 1 หมายความว่า ในเดือนนั้นมีจำนวนผู้ป่วยในต่ำกว่าค่าเฉลี่ยรายเดือน เช่น 0.89 = ต่ำกว่าค่าเฉลี่ยร้อยละ 11

ตารางที่ 1 ดัชนีฤดูกาลจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอด ของโรงพยาบาลนครพิงค์ จังหวัดเชียงใหม่

เดือน	วิธีค่าเฉลี่ยอย่างง่าย	วิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่
ม.ค.	1.01	0.93
ก.พ.	0.99	1.03
มี.ค.	1.03	0.91
เม.ย.	0.89	0.96
พ.ค.	0.96	1.00
มิ.ย.	0.99	1.01
ก.ค.	0.99	1.04
ส.ค.	1.03	1.04
ก.ย.	1.03	1.03
ต.ค.	1.04	1.03
พ.ย.	1.03	1.02
ธ.ค.	1.02	1.00
รวม	12.00	12.00

แบบจำลองจากทฤษฎีระบบเกรย์

ค่าพยากรณ์ปี พ.ศ. 2566 จากแบบจำลองตามทฤษฎีระบบเกรย์คำนวณจากข้อมูลจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอด พ.ศ. 2558 - 2565 ดังแสดงในตารางที่ 2 พบว่า GM(1,1) EPC มีค่า MAPE คือ 1.18 ต่ำที่สุดจึงเลือกนำไปแจกแจงจำนวนผู้ป่วยในรายเดือนด้วยดัชนีฤดูกาลต่อไป

การพยากรณ์จำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอด พ.ศ. 2566 รายเดือน

วิธีที่ 1 CDRMA คำนวณด้วยโปรแกรม Excel

วิธีที่ 2 SARIMA คำนวณด้วยโปรแกรม R package Forecast ได้แบบจำลองที่เหมาะสมที่สุดคือ ARIMA$(1,1,1)(0,1,1)_{12}$

วิธีที่ 3 GSA ใช้ GM(1,1) EPC คำนวณด้วยโปรแกรม Excel

วิธีที่ 4 GRMA คำนวณด้วยโปรแกรม Excel ดังแสดงในตารางที่ 3

ตารางที่ 2 การพยากรณ์จำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอดของโรงพยาบาลนครพิงค์ จังหวัดเชียงใหม่ พ.ศ. 2566 ด้วยแบบจำลองทฤษฎีเกรย์

ปี พ.ศ.	ค่าพยากรณ์				
	ค่าจริง	GM(1,1)	GM(1,1)EPC	VM	VMI
2558	49,434	49,434	49,434	49,434	49,434
2559	49,088	49,864	50,050	25,559	27,137
2560	50,859	50,942	49,984	35,743	38,252
2561	50,549	52,043	51,164	46,905	49,956
2562	55,741	53,169	55,508	56,672	59,116
2563	56,742	54,318	56,546	62,018	62,313
2564	53,548	55,493	54,133	60,928	58,165
2565	56,000	56,693	55,141	53,833	48,448
MAPE %	0	2.64	1.18	16.21	15.52
ค่าพยากรณ์ 2566	-	57,918	58,106	43,285	36,668
ร้อยละ + เพิ่ม/- ลดจากปี 2565		3.43	3.76	-22.70	-34.52

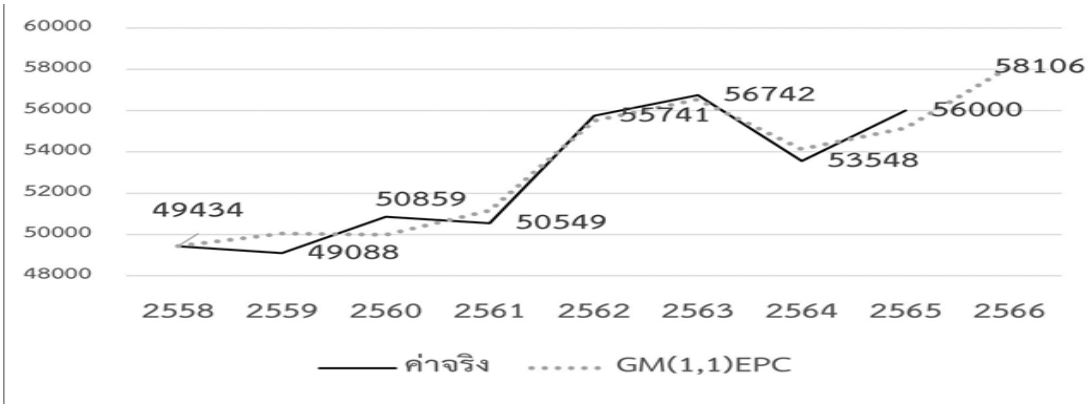
ตารางที่ 3 ค่าพยากรณ์รายเดือนจำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอดของโรงพยาบาลแห่งหนึ่งในจังหวัดเชียงใหม่ พ.ศ. 2566 ทั้ง 4 วิธี

เดือน	CDRMA	GSA	GRMA	SARIMA
ม.ค.	4,250	4,875	4,492	4,164
ก.พ.	4,733	4,800	4,997	4,570
มี.ค.	4,166	4,972	4,394	3,654
เม.ย.	4,433	4,295	4,671	3,631
พ.ค.	4,581	4,625	4,821	4,009
มิ.ย.	4,633	4,781	4,871	4,146
ก.ค.	4,794	4,799	5,035	4,208
ส.ค.	4,788	4,995	5,023	4,211
ก.ย.	4,778	4,972	5,007	4,332
ต.ค.	4,776	5,039	5,000	4,797
พ.ย.	4,722	4,996	4,939	4,522
ธ.ค.	4,647	4,956	4,855	4,579
รวม	55,300	58,106	58,106	50,823
ร้อยละ + เพิ่ม/- ลด จาก พ.ศ. 2565 มีผู้ป่วย 56,000 คน	-1.25	3.76	3.76	-9.24
MAPE 96 เดือน พ.ศ. 2558-2565	5.39	-	-	4.44
9 เดือน พ.ศ. 2566	5.13	4.05	4.88	14.90

การพิจารณาค่าพยากรณ์รายปี เมื่อพิจารณาจำนวนผู้ป่วยในรวมต่อปี พบว่า มีอัตราการเพิ่มสูงสุดใน พ.ศ. 2562 ถึงร้อยละ 10.62 จากปีก่อนหน้า และอัตราการลดลงต่ำสุดใน พ.ศ. 2564 ถึงร้อยละ 5.63 จากปีก่อนหน้า จากข้อมูลในอดีต 8 ปี พ.ศ. 2558 – 2565 จำนวนผู้ป่วย 9 เดือนแรก และ 3 เดือนสุดท้ายของปี เท่ากับร้อยละ 74.20 และ 25.80 ตามลำดับ เมื่อนำมาประมาณการจากจำนวนผู้ป่วยรวม 9 เดือน ใน พ.ศ. 2566 จำนวน 43,109 คน คิดเป็นร้อยละ 74.20 แล้วประมาณการผู้ป่วยใน พ.ศ. 2566 จะมีจำนวนถึง 58,098 คน เพิ่มจาก พ.ศ. 2565 ร้อยละ 3.76 ขณะที่ค่าเฉลี่ยรายเดือนใน พ.ศ. 2566 เพิ่มขึ้นจาก พ.ศ. 2565 เพียงร้อยละ 2.64 ซึ่งมีจำนวนผู้ป่วยตรงกับ GM(1,1)EPC ที่มีผู้ป่วย 58,106 คน MAPE ในอดีต 8 ปี เพียงร้อยละ 1.18 เท่านั้น (ตารางที่ 2)

การพิจารณาความแม่นยำจากค่าพยากรณ์รายเดือนในอดีตระยะยาวพบว่า SARIMA มีค่า MAPE 4.44 ต่ำกว่า DCRMA แต่เมื่อเทียบกับข้อมูลจริงเดือน มกราคม ถึงกันยายน 2566 ค่า MAPE 9 เดือน พบว่า GSA มีค่า MAPE ร้อยละ 4.05 ต่ำที่สุด รองลงมาคือ GRMA มีค่า MAPE ร้อยละ 4.88 ดังแสดงในตารางที่ 3 และเมื่อพิจารณาจากจำนวนผู้ป่วยรายปี ข้อมูล พ.ศ. 2558 – 2565 มีแนวโน้มเพิ่มขึ้น ดังแสดงในภาพที่ 3

ภาพที่ 3 การพยากรณ์จำนวนผู้ป่วยในในกลุ่มโรคฝุ่นจับปอด พ.ศ. 2566



วิจารณ์

การนำค่าพยากรณ์ไปใช้ เมื่อพิจารณาจากจำนวนผู้ป่วยในรายปี ข้อมูล พ.ศ. 2558 – 2565 มีแนวโน้มเพิ่มขึ้น ดังแสดงในภาพที่ 3 การเลือก GM(1,1)EPC ที่มีผู้ป่วย 58,106 คน MAPE ในอดีต 8 ปี เพียงร้อยละ 1.18 ถ้าน้อยกว่า 10 มีความแม่นยำสูง[27] (ตารางที่ 2) แล้วนำมาแจกแจงรายเดือนด้วยดัชนีฤดูกาลวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ เนื่องจากข้อมูลมีทั้งองค์ประกอบฤดูกาลและแนวโน้ม ดังแสดงค่าพยากรณ์ GRMA ในตารางที่ 4 ซึ่งมีค่า MAPE 9 เดือนสูงกว่าวิธี GSA เล็กน้อย

การตรวจสอบความแม่นยำของการพยากรณ์จำนวนผู้ป่วยใน พ.ศ. 2566 กับรายงานสถิติสาธารณสุขของจังหวัดเชียงใหม่ รายงานที่ 178 ที่พบว่า มีจำนวนผู้ป่วยใน 56,937 ราย ซึ่งวิธี GRMA ที่พยากรณ์ด้วย GM(1,1)EPC สูงกว่าจำนวนผู้ป่วยตามรายงานสถิติสาธารณสุขเพียงร้อยละ 2.05 กล่าวได้ว่า วิธีการแบบผสมเป็นการพยากรณ์ที่ความแม่นยำสูง

ผลการศึกษาพบว่า แบบจำลอง SARIMA หรือ ARIMA$(1,1,1)(0,1,1)^{(12)}$ ในการพยากรณ์ย้อนหลัง 96 เดือน โดยมีค่า MAPE ร้อยละ 4.44 ซึ่งต่ำกว่าแบบจำลอง CDRMA ที่มี MAPE ร้อยละ 5.39 สอดคล้องกับการศึกษาของ Wang และคณะ[7] ที่รายงานว่าแบบจำลอง

ARIMA มีประสิทธิภาพสูงในการพยากรณ์แนวโน้มการระบาดของโรค COVID-19 เนื่องจากสามารถจัดการกับข้อมูลอนุกรมเวลาที่มีฤดูกาลและแนวโน้มได้ดี นอกจากนี้ Ceylan[9] ยังพบว่าแบบจำลอง ARIMA มีความแม่นยำสูงในการพยากรณ์ความชุกของโรค COVID-19 ในอิตาลี สเปน และฝรั่งเศส อย่างไรก็ตาม การศึกษาในประเทศจีนโดย Lou และคณะ[5,28] เปรียบเทียบแบบจำลอง ARIMA, DNN และ LSTM ในการพยากรณ์โรคฝุ่นจับปอดจากการประกอบอาชีพในเมืองเทียนจิน กลับพบว่าแบบจำลอง DNN และ LSTM ให้ผลการพยากรณ์ที่แม่นยำกว่า ARIMA โดยมีค่า RMSE และ MAE ต่ำกว่า ซึ่งแตกต่างจากผลการศึกษาครั้งนี้ที่พบว่า ARIMA ยังคงให้ความแม่นยำสูงและมีความเหมาะสมกับการนำไปใช้ เนื่องจากมีโครงสร้างที่ไม่ซับซ้อนและสามารถตีความได้ง่าย เมื่อพิจารณาการพยากรณ์แบบผสมด้วยทฤษฎีระบบ-เกรย์ พบว่า แบบจำลอง GM(1,1) EPC มีความแม่นยำสูงที่สุดในการพยากรณ์รายปี โดยมี MAPE ร้อยละ 1.18 เมื่อนำมาแจกแจงเป็นรายเดือนด้วยดัชนีฤดูกาล พบว่า วิธี GSA และ GRMA มีค่า MAPE ร้อยละ 4.05 และ 4.88 ตามลำดับ ซึ่งอยู่ในเกณฑ์ความแม่นยำสูง MAPE น้อยกว่าร้อยละ 10 ตามเกณฑ์ของ Lewis[27] ผลการศึกษานี้สอดคล้องกับการทบทวนวรรณกรรมของ Zhang และคณะ[10] ที่ชี้ว่าแบบจำลองเกรย์ถูกนำไปใช้ในวงการแพทย์และสาธารณสุขอย่างกว้างขวาง และสามารถให้ผลการพยากรณ์ที่แม่นยำแม้มีข้อมูลจำนวนน้อย อย่างไรก็ตาม การศึกษาของ Zhou และคณะ[11] ในประเทศจีนที่พยากรณ์จำนวนผู้ป่วยโรคฝุ่นจับปอดรายใหม่ระหว่างปี พ.ศ. 2563-2564 โดยใช้ข้อมูล พ.ศ. 2543-2562 พบว่า แบบจำลอง GM(1,1) ให้ผลการพยากรณ์ด้อยกว่า Generalized Additive Model (GAM) เนื่องจากข้อมูลมีลักษณะโค้งขึ้นลงเป็นรูปตัว S ซึ่งแบบจำลอง GM(1,1) ซึ่งเป็นแบบจำลองเชิงเส้นไม่สามารถจับรูปแบบดังกล่าวได้ดีนัก แต่ในการศึกษาครั้งนี้ได้เลือกใช้ GM(1,1) EPC ซึ่งมีการแก้ไขข้อผิดพลาดเป็นระยะ ส่งผลให้สามารถพยากรณ์ข้อมูลที่มีความผันผวนได้ดีขึ้น สอดคล้องกับคำแนะนำของ Liu และ Lin[12] ที่ระบุว่าแบบจำลองเกรย์สามารถปรับค่าให้สอดคล้องกับรูปแบบข้อมูลที่หลากหลายได้

การเลือกใช้ดัชนีฤดูกาลวิธีอัตราส่วนต่อการเฉลี่ยเคลื่อนที่ (GRMA) สำหรับแจกแจงค่าพยากรณ์รายปีเป็นรายเดือน มีความเหมาะสมเนื่องจากข้อมูลมีทั้งองค์ประกอบฤดูกาลและแนวโน้มเพิ่มขึ้น สอดคล้องกับวิธีการที่ Vision Academy[18] และ Srivastava[19] นำเสนอ สำหรับข้อมูลรายเดือนที่มีความผันแปรตามฤดูกาลร่วมกับแนวโน้ม เมื่อตรวจสอบความแม่นยำ กับรายงานสถิติสาธารณสุขจังหวัดเชียงใหม่[3] พ.ศ. 2566 พบจำนวนผู้ป่วยจริง 56,937 ราย วิธี GRMA ให้ค่าพยากรณ์สูงกว่าค่าจริงเพียงร้อยละ 2.05 ยืนยันได้ว่าวิธีการแบบผสมนี้มีความแม่นยำสูงและสามารถนำไปประยุกต์ใช้ในการวางแผนบริการสุขภาพได้จริง

ข้อจำกัดของการศึกษานี้คือการใช้เพียงขนาดและรูปแบบของข้อมูลในอดีตในการพยากรณ์ โดยไม่ได้คำนึงถึงปัจจัยภายนอกที่มีผลต่อจำนวนผู้ป่วย เช่น การบังคับใช้มาตรการทางกฎหมาย การเฝ้าระวัง การลงทุนของผู้ประกอบการ และความร่วมมือในการใช้เครื่องป้องกันต่าง ๆ ซึ่งผู้บริหารในพื้นที่ย่อมต้องนำมาพิจารณาร่วมด้วย(29) การศึกษาในอนาคตควรพิจารณานำเทคนิคการพยากรณ์แบบอื่น เช่น การพยากรณ์เชิงคุณภาพ ที่คำนึงถึงความคิดเห็นของผู้เชี่ยวชาญ หรือการพัฒนาแบบจำลองที่สามารถรวมปัจจัยภายนอกเหล่านี้เข้าไปในการวิเคราะห์เพื่อเพิ่มความแม่นยำในการพยากรณ์ต่อไป

เอกสารอ้างอิง

1. Qi XM, Luo Y, Song MY, Liu Y, Shu T, Liu Y, et al. Pneumoconiosis: current status and future prospects. Chin Med J 2021;134(8):898-907.
2. วิวัฒน์ เอกบูรณะวัฒน์. คู่มือการลงรหัส ICD-10 สำหรับโรคจากการประกอบอาชีพ. ชลบุรี: สัมมาอาชีวะ; 2554.
3. ทีมพัฒนาสารสนเทศ เขตบริการสุขภาพที่ 1. รายงานอัตราป่วยรายโรค 178 ร้อยละการป่วยด้วยกลุ่มโรค โรคฝุ่นสะสม

ในปอด 2557-2566 [อินเทอร์เน็ต]. 2567 [สืบค้นเมื่อ 30 ต.ค. 2567]. แหล่งข้อมูล: https://cmi.ciorh1.com/web/index.php?r=report%2Fpdx&co_thip_new=178&chwcode=50&year=2023

4. กระทรวงสาธารณสุข HDC v4.0. อัตราป่วยโรคระบบทางเดินหายใจ เขตสุขภาพที่ 1 จังหวัดเชียงใหม่ โรงพยาบาลนครพิงค์ ปี 2556 ถึง 2566 [อินเทอร์เน็ต]. 2567 [สืบค้นเมื่อ 30 ต.ค. 2567]. แหล่งข้อมูล:https://hdcservice.moph.go.th/hdc/reports/report.php?cat_id=f16421e617aed29602f9f09d951cce68&id=4408d6d11a1efac662b6fd9e83594cea
5. Lou HR, Wang X, Gao Y, Zeng Q. Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China. BMC Public Health 2022;22(1):2167.
6. Li Z, Li Y. A comparative study on the prediction of the BP artificial neural network model and the ARIMA model in the incidence of AIDS. BMC Med Inform Decis Mak 2020;20(1):143.
7. Wang Y, Xu C, Yao S, Zhao Y. Forecasting the epidemiological trends of COVID-19 prevalence and mortality using the advanced alpha-Sutte Indicator. Epidemiol Infect 2020;148:e236.
8. Adeyinka DA, Muhajarine N. Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models. BMC Med Res Methodol 2020;20(1): 292.
9. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 2020;729:138817.
10. Zhang L, Hong T, Miao H. Grey system analysis in the field of medicine and health. Grey Systems 2019;9(2): 251-8.
11. Zhou D, Zhu D, Li N, Han B. Exploration of Three Incidence Trend Prediction Models Based on the Number of Diagnosed Pneumoconiosis Cases in China From 2000 to 2019. J Occup Environ Med 2021;63(7):e440-44.
12. Liu S, Lin Y. Grey systems theory and application. Berlin: Springer; 2010.
13. โรงพยาบาลนครพิงค์. ประวัติโรงพยาบาล [อินเทอร์เน็ต]. 2567 [สืบค้นเมื่อ 30 ต.ค. 2567]. แหล่งข้อมูล: https://www.nkp-hospital.go.th/th/history.php
14. Microsoft. Support>>Excel functions (by category) [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://support.microsoft.com/en-us/office/excel-functions-by-category-5f91f4e9-7b42-46d2-9bd1-63f26a86c0eb
15. Excel Easy. Line chart [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://www.excel-easy.com/examples/trendline.html
16. The R Foundation. The R project for statistical computing [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://www.r-project.org/
17. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, Kuroptev K, et al. Methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://cran.r-project.org/web/packages/forecast/index.html
18. Vision Academy. Measurement of seasonal indices using ratio to moving average method for monthly data [Internet]. 2021 [cited 2023 Oct 20]. Available from: https://www.youtube.com/watch?v=dqGUvOWT_14
19. Srivastava M. University of Lucknow - methods of time series [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://www.lkouniv.ac.in/site/writereaddata/siteContent/202003291612342405mukeshsrivastava_businessstastics3.pdf

20. Shweta. Towards data science – part 1: average and smoothing models [Internet]. 2021 [cited 2023 Oct 20]. Available from: https://towardsdatascience.com/introduction-to-time-series-forecasting-part-1-average-and-smoothing-models-a739d832315

21. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. 2nd edition. Melbourne: OTexts; 2018.

22. Shweta. Towards data science – introduction to time series forecasting — part 2 (ARIMA Models) [Internet]. 2021 [cited 2023 Oct 20]. Available from: https://towardsdatascience.com/introduction-to-time-series-forecasting-part-2-arima-models-9f47bf0f476b

23. CRAN. Package forecast. forecast: forecasting functions for time series and linear models [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://cran.r-project.org/web/packages/forecast/index.html

24. Cobb B. Deseasonalized data with a centered moving average [Internet]. 2020 [cited 2023 Oct 20]. Available from: https://www.youtube.com/watch?v=lCGw53-tlRw

25. Emmanuel J. Calculating seasonal indexes (without trend) in Excel | Forecasting [Internet]. 2021 [cited 2023 Oct 20]. Available from: https://www.youtube.com/watch?v=5foKJgA7mWw

26. Andrés D. Machine learning pills: error metrics for time series forecasting [Internet]. 2023 [cited 2023 Oct 20]. Available from: https://mlpills.dev/time-series/error-metrics-for-time-series-forecasting/

27. Lewis CD. Industrial and business forecasting methods. London: Butterworths; 1982.

28. Indeed Editorial Team. Quantitative forecasting vs. qualitative forecasting [Internet]. 2022 [cited 2023 Oct 20]. Available from: https://www.indeed.com/career-advice/career-development/quantitative-vs-qualitative-forecasting-pros-and-cons

28. Lou J, Li Y, Wang H, Zhang X, Chen Y. Forecasting the burden of pneumoconiosis using ARIMA, deep neural networks and LSTM models in China. BMC Public Health 2022;22:1456.

29. Indeed Editorial Team. Quantitative forecasting vs. qualitative forecasting [Internet]. 2022 [cited 2023 Oct 20]. Available from: https://www.indeed.com/career-advice/career-development/quantitative-vs-qualitative-forecasting-pros- and-cons

Forecasting the Number of Pneumoconiosis Inpatients: a Case of Hospital in Chiang Mai Province

Priracha Suthon, M.Sc.*; Thapakorn Traiyawipak, M.Sc.*; Vadhana Jayathavaj, Ph.D.**

** Faculty of Science and Technology, Chiang Mai Rajabhat University, Chiang Mai Province;*

*** Faculty of Allied Health Sciences, Pathumthani University, Pathum Thani Province, Thailand*

Journal of Health Science of Thailand 2026;35(2):234-45.

Corresponding author: Priracha Suthon, Email: priracha_sut@g.cmru.ac.th

Abstract: From fiscal year 2015 to September 2023, a hospital in Chiang Mai Province recorded monthly statistics of over 4,000 pneumoconiosis inpatients. The time series data revealed seasonal patterns and an upward trend. This study aimed to forecast monthly inpatient cases using three approaches: the Classical Decomposition-Multiplicative model with a seasonal index derived from the Ratio-to-Moving Average method (classical decomposition ratio to moving average: CDRMA), the Box and Jenkins method (seasonal autoregressive integrated moving average: SARIMA), and a hybrid model based on Grey System Theory. The forecasting used monthly inpatient data collected from the CMI reporting system, developed by the Chiang Mai Provincial Public Health Office, covering fiscal years 2014-2023. Among the models tested, ARIMA(1,1,1)$(0,1,1)_{12}$ was the most suitable, with a 96-month mean absolute percentage error (MAPE) of 4.44%, lower than the 5.39% from the CDRMA model. The hybrid Grey model, GM(1,1) with error periodic correction (EPC), achieved a MAPE of 1.18% when annual forecasts were distributed into monthly values for 2023. For the same year, the GRMA and GSA methods yielded 9-month MAPE values of 4.88% and 4.05%, respectively. Due to the seasonal and increasing nature of the data, the GRMA model was selected. Forecast accuracy was validated using the Public Health Statistics Report of Chiang Mai Province (Report No. 178), published in early 2024, which recorded 56,937 actual cases. The GRMA method, using GM(1,1)EPC, produced results that were 2.05% higher than the actual figures.

Keywords: **forecasting; inpatients; pneumoconiosis**