

Forecasting Medical Time Series Data Using the ARIMA Model

Wareeya Vongspanich

Department of Anesthesiology, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Siriraj Medical Bulletin 2021;14(4): 38-49

Abstract

Medical time series data forecasting with the Box-Jenkins method (the ARIMA model) is a quantitative prediction method with high accuracy and precision. It is a simple method, suitable for stationary time series data with constant mean and variance, the forecasters do not have to define a model before analyzing. During data analysis, the model is determined according to the time series's structure. There are only four steps to implement: 1) Model Identification 2) Parameter Estimation 3) Diagnostic Checking and 4) Forecasting time series data. The IBM SPSS software also provides a model command wizard with Expert Modeler. The program will select the most appropriate model for the data to make it easier for the forecasters to select the model, but the forecasters should reconsider the statistical significance of selecting the most accurate model again. To obtain a model suitable for that time series data this creates the least forecast error for forecasting future time series data.

Keywords: forecast; time series data; ARIMA model; IBM SPSS program; Box-Jenkins method.

Correspondence to: Wareeya Vongspanich

E-mail: wareeya.von@mahidol.edu

ORCID: <https://orcid.org/0000-0001-9998-5812>

Received: 5 May 2021

Revised: 9 August 2021

Accepted: 13 August 2021

<http://dx.doi.org/10.33192/smb.v14i4.249858>

การพยากรณ์ข้อมูลอนุกรมเวลา ทางการแพทย์ด้วยตัวแบบอาร์มา

วริยา วงศ์พานิช

ภาควิชาวิสัญญีวิทยา คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล กรุงเทพมหานคร

บทคัดย่อ

การพยากรณ์ข้อมูลอนุกรมเวลาทางการแพทย์ด้วยวิธีบ็อกซ์เจนกินส์ (ตัวแบบอาร์มา) เป็นวิธีการพยากรณ์เชิงปริมาณที่มีความถูกต้องและความแม่นยำสูง เป็นวิธีที่ง่าย เหมาะกับข้อมูลอนุกรมเวลาที่มีคุณลักษณะคงที่ ซึ่งจะมีค่าเฉลี่ยและความแปรปรวนคงที่ ผู้พยากรณ์ไม่ต้องกำหนดตัวแบบขึ้นมาก่อนทำการวิเคราะห์ ในระหว่างการวิเคราะห์ข้อมูล ตัวแบบจะถูกกำหนดขึ้นตามโครงสร้างของชุดอนุกรมเวลานั้น มีขั้นตอนการดำเนินการเพียง 4 ขั้นตอน คือ 1) การกำหนดตัวแบบของการพยากรณ์ 2) การประมาณค่าพารามิเตอร์ 3) การตรวจสอบความเหมาะสมของตัวแบบ และ 4) การพยากรณ์ข้อมูลอนุกรมเวลา โปรแกรมสำเร็จรูป ไอบีเอ็ม เอสพีเอสเอส ยังมีตัวช่วยคำสั่งสร้างตัวแบบด้วยผู้เชี่ยวชาญในการกำหนดตัวแบบของการพยากรณ์ โดยโปรแกรมจะเลือกตัวแบบที่เหมาะสมกับข้อมูลที่สุ่มมาให้ทำให้ผู้พยากรณ์สามารถเลือกตัวแบบได้สะดวกขึ้น แต่ผู้พยากรณ์ควรพิจารณาเลือกตัวแบบที่ได้มาให้ความถูกต้องมากที่สุดอย่างมีนัยสำคัญทางสถิติอีกครั้ง เพื่อจะได้ตัวแบบที่เหมาะสมกับข้อมูลอนุกรมเวลานั้นๆ ทำให้เกิดความคลาดเคลื่อนในการพยากรณ์น้อยที่สุดสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาในอนาคต

คำสำคัญ: การพยากรณ์; ข้อมูลอนุกรมเวลา; ตัวแบบอาร์มา; โปรแกรมไอบีเอ็ม เอสพีเอสเอส; วิธีบ็อกซ์เจนกินส์

บทนำ

ข้อมูลอนุกรมเวลาทางการแพทย์มีลักษณะที่มีความหลากหลาย เช่นโรคที่เกิดขึ้นตามฤดูกาล หรือโรคที่เกิดขึ้นตลอดทั้งปี การเลือกใช้เทคนิคการพยากรณ์ที่มีประสิทธิภาพจึงเป็นสิ่งสำคัญที่จะได้ตัวแบบที่เหมาะสมกับข้อมูลนั้นๆ มาพยากรณ์ให้มีความคลาดเคลื่อนน้อยที่สุด เพื่อให้ทราบจำนวนผู้ป่วยในอนาคตที่เกิดขึ้นจากโรคและภัยสุขภาพที่ใกล้เคียงกับโอกาสที่จะเกิดขึ้นจริงในอนาคต ในการเตรียมการป้องกันล่วงหน้าให้มีการเกิดน้อยลงหรือไม่เกิดขึ้นเลย ลดการสูญเสียทั้งชีวิต ทรัพย์สินและทรัพยากรต่างๆ มีหน่วยงานราชการได้ทำรายงานข้อมูลอนุกรมเวลา เช่น กรมควบคุมโรค กระทรวงสาธารณสุข มีการพยากรณ์โรค 7 โรคที่สำคัญ ได้แก่ โรคไข้เลือดออก โรคไข้หวัดใหญ่ โรคมือ เท้า ปาก โรคตาแดง โรคอาหารเป็นพิษ โรคไข้ฉี่หนู และโรคเมลิออยโดสิส ด้านภัยสุขภาพที่สำคัญ

มี 2 เรื่อง ได้แก่ การบาดเจ็บจากอุบัติเหตุจราจรในช่วงเทศกาลปีใหม่และสงกรานต์ และการป่วยด้วยโรคทางเดินหายใจในช่วงปัญหาหมอกควันภาคเหนือตอนบน¹ และสำนักโรคระบาด กรุงเทพมหานคร ได้พยากรณ์จำนวนผู้ป่วยที่เป็นโรคที่ต้องเฝ้าระวังทางระบาดวิทยา ได้แก่ โรคอุจจาระร่วงเฉียบพลัน โรคไข้เลือดออก โรคไข้หวัดใหญ่ โรคตาแดง โรคปอดอักเสบ โรคมือเท้าและปาก โรคสุกใส โรคอาหารเป็นพิษ โรคแผลในปาก โรควันโรคปอด เป็นต้น บทความนี้ได้นำข้อมูลอนุกรมเวลาของการป่วยด้วยโรคอาหารเป็นพิษ (food poisoning) ของจังหวัดกรุงเทพมหานคร มาเป็นตัวอย่างในการวิเคราะห์ด้วยเทคนิคการพยากรณ์ เนื่องจากเล็งเห็นว่า โรคนี้เป็นโรคที่ใกล้ตัวเพราะเกี่ยวข้องกับโภชนาการการกินในชีวิตประจำวัน รวมทั้งกรมควบคุมโรค กระทรวงสาธารณสุขและสำนักโรคระบาด กรุงเทพมหานคร ให้ความสำคัญจัดเป็นโรคที่ต้องเฝ้าระวังทางระบาดวิทยาด้วย

การพยากรณ์โดยวิธีการวิเคราะห์อนุกรมเวลามีสมมติฐานว่าประวัติศาสตร์ย่อมซ้ำรอย ดังนั้นจึงสามารถใช้ข้อมูลในอดีตเฉพาะเรื่องที่ต้องการคาดคะเนมาใช้พยากรณ์สิ่งนั้นในอนาคต ซึ่งถ้าต้องการพยากรณ์สิ่งใดก็ต้องใช้ข้อมูลในอดีตของสิ่งนั้นเท่านั้นมาใช้ในการพยากรณ์โดยไม่คำนึงถึงตัวแปรอื่นๆ ในปัจจุบันนี้ การพยากรณ์เชิงปริมาณที่ใช้กันอยู่มีด้วยกัน 8 วิธี² คือ

1. วิธีทำให้เรียบเอ็กโปเนนเชียลอย่างง่าย
2. วิธีของโฮลท์
3. วิธีของวินเตอร์
4. วิธีแยกส่วนประกอบ
5. วิธีของบ็อกซ์-เจนกินส์
6. วิธีวิเคราะห์การถดถอยแบบพหุ
7. วิธีวิเคราะห์เชิงเศรษฐกิจ และ

8. วิธีวิเคราะห์ของบ็อกซ์-เจนกินส์แบบพหุ มีโปรแกรมสำเร็จรูปมากมายในการช่วยวิเคราะห์สถิติพยากรณ์ เช่น MINITAB, SPSS/PC+, IBM SPSS, SAS, MedCalc เป็นต้น

ในปี ค.ศ. 1972 George E.P. Box และ Gwilym M. Jenkins³ ได้พัฒนาวิธีการวิเคราะห์ข้อมูลอนุกรมเวลาขึ้นโดยใช้ ตัวแบบจำลอง ARIMA (Auto Regressive Intergrated Moving Average) หรือวิธีบ็อกซ์-เจนกินส์ (Box-Jenkins mothod) เป็นการนำ Regression technique มาประยุกต์ใช้กับข้อมูลอนุกรมเวลา ต้องมีการตรวจสอบข้อมูลอนุกรมเวลา ว่ามีคุณลักษณะคงที่หรือไม่ก่อนดำเนินการวิเคราะห์

สำหรับการพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์เป็นเทคนิค

การพยากรณ์ที่มีจุดเด่น คือ

1. เป็นวิธีให้ค่าพยากรณ์ไปข้างหน้าที่ดีทั้งในช่วงระยะสั้น ระยะกลาง และระยะยาว^{2, 7}
2. เหมาะกับกรณีที่ใช้วิธีการพยากรณ์อื่นไม่ได้ผล ข้อมูลอนุกรมเวลาที่มีคุณลักษณะคงที่
3. เป็นวิธีที่ไม่มีการกำหนดตัวแบบตายตัวขึ้นก่อนทำการวิเคราะห์ ในระหว่างการวิเคราะห์ตัวแบบจะถูกกำหนดขึ้นมาเองตามโครงสร้างของอนุกรมเวลานั้นๆ ทำให้ตัวแบบที่ได้มีความคลาดเคลื่อนต่ำ
4. เป็นวิธีพยากรณ์เชิงปริมาณวิธีหนึ่งที่มีความถูกต้องและแม่นยำสูงสุดสำหรับ 2-3 ค่าแรกของการพยากรณ์^{2,3,6,7}
5. เป็นวิธีที่ง่ายต่อการเข้าใจ แม้ผู้พยากรณ์จะขาดประสบการณ์ แต่มีความเข้าใจเรื่องกราฟอัตโนมัติสัมพันธ์ก็อาจได้ตัวแบบในการพยากรณ์ที่ดีได้โดยใช้โปรแกรมสำเร็จรูป⁵

จุดด้อย คือ

1. จำนวนข้อมูลย้อนหลังที่นำมาวิเคราะห์อย่างน้อย 50 ค่า แต่โดยมากแล้วจะใช้ประมาณ 100 ค่าเป็นต้นไป การเตรียมข้อมูล ยิ่งมีข้อมูลมากเท่าไร ยิ่งมีความแม่นยำในการพยากรณ์มากเท่านั้น สำหรับข้อมูลที่มีอิทธิพลของฤดูกาลต้องใช้จำนวนข้อมูลย้อนหลัง 5 รอบของฤดูกาล (5*S) ซึ่งต้องมีค่าเฉลี่ยคงที่ ฤดูกาลคงที่ มีความแปรปรวนเล็กน้อย^{2,7}
2. ไม่เหมาะกับข้อมูลอนุกรมเวลาที่มีรูปแบบของอิทธิพลแนวโน้ม วัฏจักร และฤดูกาลที่เด่นชัด ควรเลือกใช้เทคนิคการพยากรณ์วิธีอื่นให้เหมาะสม

กับข้อมูลอนุกรมเวลานั้นๆ เพื่อจะให้ความคลาดเคลื่อนน้อยที่สุด

3. การคำนวณด้วยตนเองด้วยมือมีความยุ่งยาก ซับซ้อน ต้องใช้คอมพิวเตอร์เพื่อเพิ่มประสิทธิภาพการทำงานให้ง่ายขึ้น

นิยามคำศัพท์

นิยามศัพท์เฉพาะมีดังนี้

ข้อมูลอนุกรมเวลา คือ ข้อมูลที่มีการเปลี่ยนแปลงเมื่อเวลาเปลี่ยนไป มักจัดเก็บเป็นข้อมูลตามคาบเวลาที่แน่นอน เช่น รายปี ครึ่งปี รายไตรมาส รายเดือน รายวัน เป็นต้น

ข้อมูลอนุกรมเวลา ประกอบด้วยองค์ประกอบ 4 ส่วน⁴ คือ แนวโน้ม (Trend: T), ฤดูกาล (Seasonal: S), วัฏจักร (Cycle: C) และเหตุการณ์ที่ผิดปกติหรือเหตุการณ์ความไม่แน่นอน (Irregular: I)

ข้อมูลอนุกรมเวลาทางการแพทย์ คือ ข้อมูลเชิงปริมาณทางการแพทย์ สาธารณสุขมีการเก็บรวบรวมตามคาบของเวลาอย่างต่อเนื่อง หรือข้อมูลเชิงปริมาณทางการแพทย์สาธารณสุขที่ขึ้นอยู่กับปัจจัยเวลา

ARIMA model คือ ตัวแบบ ARIMA (p, d, q) มีส่วนประกอบที่สำคัญ 3 ส่วน⁴ ได้แก่ AutoRegressive AR: (p), Intergrated I: (d) และ Moving Average MA: (q)

SARIMA model คือ ตัวแบบอนุกรมเวลาที่มีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง พิจารณาพารามิเตอร์ Seasonal I(D), Seasonal AR(P) และ Seasonal MA(Q) เข้าไปในตัวแบบ ARIMA(p,d,q)x(P,D,Q)⁵

การพยากรณ์ คือ การประมาณการ หรือ การคาดคะเนว่าอะไรจะเกิดขึ้นช่วงเวลาในอนาคต

วิธีการพยากรณ์ แบ่งตามลักษณะข้อมูล มี 2 ประเภท⁵ คือ

1. การพยากรณ์เชิงปริมาณ อาศัยข้อมูลในอดีตเป็นหลักในการคาดคะเนและประมาณการสิ่งใดที่จะเกิดในช่วงเวลาในอนาคต
2. การพยากรณ์เชิงคุณภาพ ไม่อาศัยข้อมูลในอดีตเป็นหลัก แต่ใช้ความรู้สึกลำดับและประสบการณ์ต่างๆ ที่ผ่านมา

แบ่งตามระยะเวลาในการพยากรณ์ มี 4 ระยะ⁶ คือ

1. การพยากรณ์ 1 หน่วยเวลาล่วงหน้า (intermediate-term forecasting)
2. การพยากรณ์ช่วงสั้น (short-term forecasting)
3. การพยากรณ์ช่วงกลาง (middle-term forecasting)
4. การพยากรณ์ระยะยาว (long-term forecasting)

ข้อตกลงเบื้องต้นของการวิเคราะห์ด้วยวิธีบ็อกซ์-เจนกินส์

การทดสอบอนุกรมเวลา เป็นสิ่งที่ควรกระทำก่อนที่จะนำข้อมูลอนุกรมเวลาใช้ในการวิเคราะห์ โดยเฉพาะเงื่อนไขความคงที่ของอนุกรมเวลา (stationary) ซึ่งเป็นเงื่อนไขที่สำคัญในการนำข้อมูลอนุกรมเวลาไปใช้ ดังนั้นถ้าหากอนุกรมเวลาที่ไม่คงที่ที่จะต้องทำให้อนุกรมเวลา ดังกล่าวคงที่ก่อน

ข้อมูลที่มีลักษณะคงที่ (stationary) หมายถึง ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย (mean) และ ความแปรปรวน (variance) เท่ากันตลอด

ระยะเวลาที่ศึกษา และข้อมูลไม่มีการเปลี่ยนแปลงตามฤดูกาล ส่วนข้อมูลที่มีลักษณะไม่คงที่ (non-stationary) หมายถึงข้อมูลอนุกรมที่มีค่าเฉลี่ย (mean) และความแปรปรวน (variance) ไม่เท่ากันตลอดระยะเวลาที่ศึกษา หรือข้อมูลแสดงการเปลี่ยนแปลงตามฤดูกาล

การปรับข้อมูลให้เป็น Stationary^{2,3,5}

1. ถ้าอนุกรมเวลา (Y_t) มีแนวโน้ม ต้องปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม โดยใช้วิธี Regular Differencing (ใช้สัญลักษณ์ d) กรณีที่หาความแตกต่างแค่ลำดับเดียว ($d = 1$)
2. ถ้าอนุกรมเวลา (Y_t) มีความเป็นฤดูกาล ต้องปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล โดยใช้วิธี Seasonal Differencing (ใช้สัญลักษณ์ D) กรณีที่หาความแตกต่างของฤดูกาลแค่ลำดับเดียว ($D = 1$)
3. ถ้าอนุกรมเวลา (Y_t) มีแนวโน้มและมีความเป็นฤดูกาล ต้องปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีทั้งแนวโน้มและฤดูกาล โดยใช้วิธี Regular and Seasonal Differencing กรณีที่หาความแตกต่างของทั้งแนวโน้มและฤดูกาลแค่ลำดับเดียว ($d = 1$ และ $D = 1$)
4. กรณีที่ความแปรปรวนของอนุกรมเวลาไม่คงที่ ($\text{Var}(y_t)$ ไม่คงที่ในแต่ละเวลา t) จะต้องการปรับอนุกรมเวลา (Y_t) ให้เป็นอนุกรมเวลาใหม่ โดยการทำให้ Logarithmic transformation หรือ Square root transformation

การตรวจสอบความเหมาะสมของรูปแบบหรือสมการการพยากรณ์ สามารถศึกษาเพิ่มเติมได้จาก (Box et al., 1994)³

อนุกรมเวลา (Y_t) ที่ศึกษาครั้งนี้เป็นอนุกรมเวลาตัวแบบ ARIMA (p, d, q) (P, D, Q)⁵ มีรูปแบบคือ

$$(1-B)^d(1-B)^D(1-\phi_1B-\phi_2B^2-\dots-\phi_pB^p)(1-\Phi_1B^S-\Phi_2B^{2S}-\dots-\Phi_{PS}B^{PS})Y_t = (1-\Theta_1B-\Theta_2B^2-\dots-\Theta_qB^q)(1-\Psi_1B^S-\Psi_2B^{2S}-\dots-\Psi_{QS}B^{QS})\epsilon_t$$

โดยที่

Y_t คือ ค่าสังเกตของอนุกรมเวลา ณ เวลา t (ที่เป็นสแตชันนารี)

ϕ, Φ, Θ, Ψ คือ พารามิเตอร์ในตัวแบบที่ต้องการประมาณค่า

B คือ ตัวดำเนินการถอยหลังเวลา นั่นคือ $B^m Y_t = Y_{t-m}$

d คือ จำนวนครั้งของการทำผลต่างแบบมีแนวโน้มเพื่อให้อนุกรมเวลา (Y_t) อยู่ในสภาวะคงที่

D คือ จำนวนครั้งของการทำผลต่างแบบมีฤดูกาลเพื่อให้อนุกรมเวลา (Y_t) อยู่ในสภาวะคงที่

p และ P คือ ลำดับของตัวแบบการถดถอยภายในฤดูกาล (p) และระหว่างฤดูกาล (P)

q และ Q คือ ลำดับของตัวแบบค่าเฉลี่ยเคลื่อนที่ภายในฤดูกาล (q) และระหว่างฤดูกาล (Q)

ϵ_t คือ ความคลาดเคลื่อนสุ่มที่มีการแจกแจงแบบปกติ ณ เวลา t ซึ่งมีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ เท่ากับ σ^2_{ϵ}

และเป็นอิสระต่อกัน

S คือ จำนวนฤดูกาลต่อปี ในที่นี้ $S = 12$

การวินิจฉัยตัวแบบการพยากรณ์โดยวิธีบ็อกซ์เจนกินส์^{5,6}

การประเมินค่าความแม่นยำของการพยากรณ์ ควรเลือกตรวจมากกว่า 2 รายการ สำหรับบทความนี้ใช้ค่าสถิติ เกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย และเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์เฉลี่ย เพื่อทดสอบความแม่นยำของผลการพยากรณ์ ค่า Error ยิ่งต่ำ ยิ่งดี

เกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย RMSE (Root Mean Squared Error) ใช้พิจารณาหาจุดเหมาะสมสำหรับรูปแบบหรือสมการการพยากรณ์ (optimal forecasting model) ด้วยการประมาณค่าพารามิเตอร์ของสมการที่ทำให้ค่า RMSE มีค่าน้อยที่สุด, ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย MAE (Mean Absolute Error) มีค่ายิ่งต่ำยิ่งดี ถ้าเท่ากับ 0 แปลว่าโมเดลทำนายค่า Y ได้ถูกต้อง 100% และเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์เฉลี่ย MAPE (Mean Absolute Percentage Error) $\text{MAPE} < 10\%$ ถือว่าการพยากรณ์นั้นมีความแม่นยำมาก, $\text{MAPE} 10 - 20\%$ ถือว่าการพยากรณ์นั้นดี, $\text{MAPE} 20 - 50\%$ ถือว่าการพยากรณ์นั้นพอใช้ได้ และ $\text{MAPE} > 50\%$ ถือว่าการพยากรณ์นั้นไม่มีความแม่นยำ

ขั้นตอนของวิธีบ็อกเจนกินส์²

การวิเคราะห์อนุกรมเวลาจะต้องเขียนกราฟข้อมูลชุดนั้นก่อนเสมอ การศึกษาความเคลื่อนไหวของอนุกรมเวลาเพื่อพิจารณาเบื้องต้นว่าข้อมูลอนุกรมเวลานั้นมีลักษณะแบบใด ทำการทดสอบองค์ประกอบแนวโน้มและความแปรผันตามฤดูกาล ในกรณีที่ข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์มีคุณสมบัติไม่คงที่ (Non-stationary) จะต้องทำการแปลงข้อมูลอนุกรมเวลาดังกล่าวให้มีคุณสมบัติคงที่ก่อน

ขั้นตอนที่ 1 การกำหนดตัวแบบของการพยากรณ์ (model Identification) ซึ่งเป็นขั้นตอนการพิจารณาว่าตัวแบบใดเหมาะสมที่สุดกับข้อมูลที่นำมาวิเคราะห์ โดยพิจารณาจากกราฟอัตโนมัติสหสัมพันธ์ (Auto Correlation Function: ACF) และกราฟอัตโนมัติสหสัมพันธ์บางส่วน (Partial Auto Correlation Function: PACF) ระบุลำดับ $\text{AR}(p)$ และ $\text{MA}(q)$ โดยพิจารณา $\text{AR}(p)$ จากกราฟ PACF และ $\text{MA}(q)$ จากกราฟ ACF ภายในฤดูกาล ระบุลำดับ Seasonal $\text{AR}(P)$ และ Seasonal $\text{MA}(Q)$ สำหรับ Seasonal $\text{AR}(P)$ พิจารณาจากกราฟ PACF และ Seasonal $\text{MA}(Q)$ พิจารณาจากกราฟ ACF ระหว่างฤดูกาลโดยพิจารณา Lag ที่ 12, 24 ตามรูปแบบที่บ็อกซ์เจนกินส์กำหนดไว้ การพิจารณาระบุลำดับ $\text{AR}(p)$, $\text{MA}(q)$, $\text{SAR}(P)$ และ $\text{SMA}(Q)$ ต้องทำการทดสอบว่า Lag ใดที่มีค่าแตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติจึงจะสามารถระบุลำดับได้อย่างถูกต้อง

ตารางที่ 1 แสดงลักษณะกราฟ ACF และ PACF เพื่อระบุลำดับ AR(p), MA(q), SAR(P) และ SMA(Q)

ตัวแบบ	ลักษณะกราฟ ACF	ลักษณะกราฟ PACF
ARIMA(1,0,1) คือ AR(1), MA(1)	Lag 1 $\neq 0$ หลังจาก Lag 1 มีค่า = 0	Lag 1 $\neq 0$ หลังจาก Lag 1 มีค่า = 0
ARIMA(1,0,0) คือ AR(1), MA(0)	ลดลงเข้าใกล้ 0 อย่าง รวดเร็ว	Lag 1 $\neq 0$ หลังจาก Lag 1 มีค่า = 0
ARIMA(0,0,1) คือ AR(0), MA(1)	Lag 1 $\neq 0$ หลังจาก Lag 1 มีค่า = 0	ลดลงเข้าใกล้ 0 อย่าง รวดเร็ว
SAR(1) ¹² *กรณีฤดูกาลเป็น รายเดือน S=12	Lag 12, Lag 24,... ลดลงเข้าใกล้ 0 อย่าง รวดเร็ว	Lag 12 $\neq 0$ หลังจาก Lag 12 มีค่า = 0 (คือ Lag 24, Lag 36, Lag 48....)
SMA(1) ¹² *กรณีฤดูกาลเป็น รายเดือน S=12	Lag 12 $\neq 0$ หลังจาก Lag 12 มีค่า = 0 (คือ Lag 24, Lag 36, Lag 48....)	Lag 12, Lag 24,... ลดลงเข้าใกล้ 0 อย่าง รวดเร็ว
SARIMA(1,0,1) ¹² *กรณีฤดูกาลเป็น รายเดือน S=12	Lag 12 $\neq 0$ หลังจาก Lag 12 มีค่า = 0 (คือ Lag 24, Lag 36, Lag 48....)	Lag 12 $\neq 0$ หลังจาก Lag 12 มีค่า = 0 (คือ Lag 24, Lag 36, Lag 48....)

ที่มา: วรียา วงศ์พานิช

ขั้นตอนที่ 2 การประมาณค่าพารามิเตอร์ (parameter estimation) ของตัวแบบที่กำหนดด้วยวิธีควรจะเป็นค่าสูงสุด (maximum likelihood) ตรวจสอบตัวแบบจำลอง (ปัญหา autocorrelation) ด้วยวิธี Ljung-Box Q-statistics ดูค่า Ljung-Box test (ถ้า p value > 0.05 ให้ถือว่า error มีความเป็นอิสระกัน)

ขั้นตอนที่ 3 การตรวจสอบความเหมาะสมของตัวแบบ (diagnostic checking) อนุกรมเวลาหนึ่งๆ อาจมีมากกว่า 1 ตัวแบบ ดังนั้นควรเลือกตัวแบบที่เหมาะสมที่สุด จะต้องมีการตรวจสอบความเหมาะสม แต่ถ้าหากตัวแบบที่กำหนดไม่เหมาะสมจะต้องมีการกำหนดตัวแบบในขั้นตอนที่ 1 ใหม่

ขั้นตอนที่ 4 การพยากรณ์ (forecasting) พยากรณ์ออกไปข้างหน้าจำนวน n ช่วงเวลา สำหรับโปรแกรม IBM SPSS กระทำด้วยวิธี backcasting ก่อนที่จะนำค่าประมาณพารามิเตอร์มาสร้างสมการการพยากรณ์และทำการเปรียบเทียบประสิทธิภาพ (ความแม่นยำ) ของตัวแบบพยากรณ์ (forecast evaluation statistics) โดยใช้สถิติวัดความคลาดเคลื่อน เช่น ME, RMSE, MAE MPE และ MAPE

วัตถุประสงค์

วัตถุประสงค์ของบทความนี้เพื่อนำเสนอขั้นตอนการวิเคราะห์ข้อมูลอนุกรมเวลาทางการแพทย์ด้วยวิธีบ็อกซ์-เจนกินส์ (Box-Jenkins method) เพื่อหาตัวแบบอาร์มา (ARIMA model) ด้วยโปรแกรมสำเร็จรูป IBM SPSS โดยการพยากรณ์ข้อมูลอนุกรมเวลาการป่วยด้วยโรคอาหารเป็นพิษ (Food poisoning) ของจังหวัดกรุงเทพมหานคร

ตัวอย่างการพยากรณ์

บทความนี้ใช้ข้อมูลอนุกรมเวลาของการป่วยด้วยโรคอาหารเป็นพิษ (food poisoning) ของจังหวัดกรุงเทพมหานคร โดยใช้ข้อมูลจากเว็บไซต์ของระบบฐานข้อมูลแบบรายงาน 506 ของกรมควบคุมโรค กระทรวงสาธารณสุข ตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึง ธันวาคม พ.ศ. 2562 จำนวน 204 ค่า ได้แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึงธันวาคม พ.ศ. 2561 จำนวน 192 ค่า เพื่อใช้ในการหา model ที่เหมาะสม ชุดที่ 2 ข้อมูลเดือนมกราคม พ.ศ.2562 ถึง ธันวาคม พ.ศ. 2562 จำนวน 12 ค่า นำมาทำการพยากรณ์ค่าที่ตัดออกไป เรียกว่า วิธี backcasting หาค่าพยากรณ์ย้อนหลังเพื่อการวินิจฉัยตัวแบบตรวจสอบความแม่นยำของตัวแบบ

เริ่มต้นด้วย การเตรียมข้อมูล เปิด Data file.sav ซึ่งประกอบด้วย Food Poisoning (case) ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2546 ถึงธันวาคม พ.ศ. 2561 จำนวน 192 ค่า สำหรับสร้างตัวแบบอาร์มา เราต้องกำหนดให้โปรแกรมรู้จักข้อมูลอนุกรมเวลาก่อนด้วยการคลิกที่เมนู Data > Define dates > เลือกรูปแบบเวลาที่ตรงกับข้อมูลของเรา (เช่น ถ้าเป็นข้อมูลรายเดือน ก็คือ Years, months) และระบุเวลาที่เริ่มต้น เช่น Year = 2546, month = 1 click OK (ภาพที่ 1A) จากนั้นจะสังเกตเห็นว่ามีตัวแปรใหม่เกิดขึ้นในข้อมูลของเรา (YEAR_, MONTH_, DATE_) ดัง (ภาพที่ 1B)

ขั้นตอนที่ 1 การกำหนดตัวแบบของการพยากรณ์

(Model Identification) ก่อนอื่นให้ทำการตรวจสอบว่าข้อมูลอนุกรมเวลามีคุณสมบัติคงที่แล้วหรือไม่ โดยจะต้องเขียนกราฟของข้อมูลชุดนั้นก่อนเสมอ ตามหลักการแล้วการสรุปว่าอนุกรมเวลานั้นมีองค์ประกอบที่เป็นแนวโน้มหรือฤดูกาลหรือไม่นั้นจะต้องพิจารณาจากกราฟ ACF และ PACF ตามเงื่อนไขที่ชี้แจงไว้ในขั้นตอนที่ 1 เพื่อพิจารณาความเคลื่อนไหวของอนุกรมเวลาว่ามีองค์ประกอบของอนุกรมเวลาเป็นอะไรบ้าง หรือลักษณะของอนุกรมเวลามีค่าเฉลี่ยคงที่หรือไม่ โดยใช้โปรแกรม IBM SPSS สร้างกราฟ มีคำสั่งดังนี้

1. โดยคลิก Analyze > Forecasting > Sequence Charts
2. ในช่อง Variables ให้เลือกตัวแปรที่เราต้องการพยากรณ์ เช่น Food Poisoning (case)
3. ในช่อง Time Axis Labels ให้เลือกตัวแปร Date. Format: “MMM”
4. กด OK เพื่อดูลักษณะของกราฟ

การแปลงข้อมูลอนุกรมเวลา ในกรณีที่ข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์คุณสมบัติไม่คงที่ (Non-stationary) จะต้องทำการแปลงข้อมูลอนุกรมเวลาดังกล่าวให้มีคุณสมบัติคงที่ก่อน

1.1 โดยการคลิกเมนู Analyze > Forecasting > Create models ดัง (ภาพที่ 2A)

1.2 Variable tab (ภาพที่ 2D)

ในช่อง Dependent variable ให้เลือกตัวแปรที่ต้องการวิเคราะห์ เช่น Food Poisoning (case)

ในช่อง Independent variable ให้เลือกตัวแปรที่ต้องการวิเคราะห์ เช่น MONTH, period 12_MONTH_1

ในช่อง Method ให้เลือกวิธีการวิเคราะห์ที่จะใช้เป็น ARIMA คลิก Criteria dialog box ไปสู่ Model tab

ใน ARIMA Criteria dialog box ช่อง ARIMA order ให้ปล่อยให้ค่า AR และ MA เป็น 0 ไว้ก่อน แต่ให้เลือก Difference nonseasonal = 1 และ Transformation ในรูปแบบที่จะทำให้ข้อมูลเป็น Stationary โดยทั่วไปมักไม่ทำการหาผลต่าง (Difference) มากเกิน 2 ครั้ง ข้อมูลอนุกรมเวลาก็จะเป็น Stationary แล้ว

1.3 Model tab (ภาพที่ 2B) กำหนด $p=0$, $d=1$, $q=0$ และ $P=0$, $D=0$, $Q=0$

การ difference 1 ครั้ง เพื่อให้ดูการเปลี่ยนแปลงกราฟให้แอมพิจูดใกล้เคียงกัน ทำให้ความแม่นยำ ARIMA ดีขึ้น การ difference 1 ครั้ง เป็นการจัดการข้อมูลให้น้อยที่สุด, d เป็นการ difference TREND, D เป็นการ difference SEASONAL ให้ค่าแอมพิจูดของกราฟคงที่

1.4 Plots tab (ภาพที่ 2C)

ในช่อง Plots for individual Models ให้เลือก Residual ACF และ Residual PACF กด OK เพื่อดูลักษณะ

กราฟ ACF และ PACF หาค่าลำดับ AR (p), MA (q), Seasonal AR (P) และ Seasonal MA(Q) ที่เหมาะสมดัง (ภาพที่ 2E)

จากกราฟ ACF และ PACF ดังภาพ ที่ 3A พบว่า อนุกรมเวลายังไม่คงที่ ดังนั้นผู้วิจัยจึงแปลงข้อมูลด้วยผลต่างแนวนอนลำดับ ที่ 1 ($d=1$, $D=0$, $S=12$) ได้กราฟ ACF และ PACF ของอนุกรมเวลาที่แปลงข้อมูลแล้ว แสดงดังภาพ ที่ 3B ซึ่งพบว่า อนุกรมเวลามีลักษณะคงที่ และในภาพที่ 3C แสดงรายละเอียดค่าสัมประสิทธิ์อัตโนมัติสหสัมพันธ์และสัมประสิทธิ์อัตโนมัติสหสัมพันธ์บางส่วนของค่าคลาดเคลื่อนตกอยู่ในขอบเขต ความเชื่อมั่นร้อยละ 95 ของตัวแบบ SARIMA(1, 1, 1)(1, 0, 1)¹²

ขั้นตอนที่ 2 การประมาณค่าพารามิเตอร์ (Parameter Estimation) ทดสอบ model ที่เลือกมาได้

2.1 โดยการคลิกเมนู Analyze > Forecasting > Create models ดัง (ภาพที่ 4A)

2.2 Variable tab (ภาพที่ 4B)

ในช่อง Dependent variable ให้เลือกตัวแปรที่ต้องการวิเคราะห์ เช่น Food Poisoning (case)

ในช่อง Independent variable ให้เลือกตัวแปรที่

ต้องการวิเคราะห์ เช่น MONTH, period 12_MONTH_1

ในช่อง Method ให้เลือกวิธีการวิเคราะห์ที่จะใช้เป็น ARIMA

ใน ARIMA Criteria dialog box ช่อง ARIMA order ให้นำตัวแบบที่คิดว่าเหมาะสมมาใส่ สำหรับตัวอย่างนี้ได้ตัวแบบ ARIMA (1, 1, 1) × (1, 0, 1)¹²

2.3 Model tab (ภาพที่ 4C)

จะได้ $p=1$, $d=1$, $q=1$ และ $P=1$, $D=0$, $Q=1$

2.4 Statistics tab (ภาพที่ 4D)

เลือก fit statistics ที่สนใจ โดยทั่วไปแนะนำให้เลือก Display fit measures, Ljung Box statistic ค่าสถิติ R square, RMSE, MAPE, MAE ในช่อง Statistics for individual models ให้เลือก Parameter estimates

2.5 Plots tab (ภาพที่ 4E)

ไม่ต้องดูกราฟ ACF, PACF อีกเพราะรู้ค่าลำดับของ p d q , P D Q แล้ว

Click OK เพื่อให้ทำการวิเคราะห์ ได้ค่าสถิติความคลาดเคลื่อน R square, RMSE, MAPE, MAE, ค่า Ljung Box Q test, p-value >0.05 และตาราง ARIMA Models Parameters ดัง (ภาพที่ 4F)

ขั้นตอนที่ 3 การตรวจสอบความเหมาะสมของตัวแบบ (Diagnostic Checking) ในการพยากรณ์ การทำการพยากรณ์ค่าที่ตัดออกไป เรียกว่า วิธี backcasting มีขั้นตอนดังนี้ นำไฟล์ Data file.sav ซึ่งมีข้อมูลตัวแปร case ตั้งแต่เดือนมกราคม 2546 ถึง ธันวาคม 2562 จำนวน 204 ค่า เพิ่มตัวแปรใหม่ชื่อว่า backcast โดยตัดช่วงปี 2562 ออก มีข้อมูลตั้งแต่เดือนมกราคม 2546 ถึง ธันวาคม 2561 จำนวน 192 ค่า และหาค่าพยากรณ์ ทำนายช่วงปี 2562 จำนวน 12 ค่า

3.1 เราต้องกำหนดให้โปรแกรมรู้จักข้อมูลอนุกรมเวลาก่อนด้วยการคลิกที่เมนู Data > Define dates > เลือกรูปแบบเวลาที่ตรงกับข้อมูลของเรา (เช่น ถ้าเป็นข้อมูลรายเดือน ก็คือ Years, months) และระบุเวลาที่เริ่มต้น เช่น Year = 2546, month = 1 click OK จากนั้นจะสังเกตเห็นว่ามีตัวแปรใหม่เกิดขึ้นในข้อมูลของเรา คือ (YEAR_, MONTH_, DATE_) เมื่อเชื่อมวันเดือนใหม่แล้ว ให้ run อีกครั้งใช้ ค่า p d q P D Q เดิม $p=1$, $d=1$, $q=1$ และ $P=1$, $D=0$, $Q=1$ โดยไม่ต้องดูค่าความคลาดเคลื่อนอีก ทำการพยากรณ์ตัวแบบที่เลือกมาได้ สำหรับขั้นตอนที่ 3.2 – 3.5 สามารถดูการลงรายละเอียดในแต่ละขั้นตอนได้จากภาพที่ 4 เพื่อความเข้าใจมากยิ่งขึ้น

3.2 โดยการคลิกเมนู Analyze > Forecasting > Create models

3.3 Variable tab

ในช่อง Dependent variable ให้เลือกตัวแปรที่ต้องการวิเคราะห์ เช่น backcast

ในช่อง Independent variable ให้เลือกตัวแปรที่

ต้องการวิเคราะห์ เช่น MONTH, period 12 MONTH_1

ในช่อง Method ให้เลือกวิธีการวิเคราะห์ที่จะใช้เป็น ARIMA

ใน ARIMA Criteria dialog box ช่อง ARIMA order ให้นำตัวแบบที่คิดว่าเหมาะสมมาใส่ สำหรับตัวอย่างนี้ได้ตัวแบบ ARIMA (1, 1, 1) × (1, 0, 1)¹²

3.4 Model tab จะได้ p=1,d=1, q=1 และ P=1,D=0, Q=1

3.5 Statistics tab เลือก fit statistics ที่สนใจ โดยทั่วไปแนะนำให้เลือก Display fit measures, Ljung Box statistic ค่าสถิติ R square, RMSE, MAPE, MAE

3.6 Plots tab ในช่อง Plots for individual Models ให้เลือก Series จากนั้นเลือก Observed values, Forecasts และ Fit values

3.7 Save tab อาจเลือกให้แสดง Predicted ถ้าต้องการดูค่าพยากรณ์โดยยังไม่ต้องการให้บันทึกไว้ใน data file

3.8 Options tab ใช้ตอนที่เราได้ตัวแบบที่คิดว่าดีที่สุดแล้วและต้องการให้พยากรณ์ค่าในอนาคต โดยให้เลือกคลิก First case after end of estimation period through a specified date และระบุช่วงเวลาที่ต้องการให้พยากรณ์ไปถึง เช่น Year = 2562, Month = 12 Click OK เพื่อให้ทำการพยากรณ์ ได้ผลลัพธ์ดัง (ภาพที่ 5)

ขั้นตอนที่ 4 การพยากรณ์ (Forecasting) หลังจากการทำ back-casting เพื่อตรวจสอบการพยากรณ์ เมื่อไปดูที่ หน้าจอ Data file จะพบว่า มีตัวแปรใหม่เกิดขึ้น คือ

Predicted_... คือ ค่าที่ได้จากการพยากรณ์ ซึ่งจะประกอบด้วย

1. ค่าพยากรณ์ของข้อมูล record แรกจนถึงช่วงเวลาที่เรามีข้อมูลย้อนสุดท้าย (Fit values) เช่น ข้อมูลปี พ.ศ. 2546 - 2561

2. ค่าพยากรณ์เท่าที่เราต้องการให้พยากรณ์ทำนายไปข้างหน้า (Forecasts) เช่น ข้อมูลปี พ.ศ.2562

NResidual_... คือ ค่าความคลาดเคลื่อน (Error) หรือ (ข้อมูล

จริง – ค่าทำนาย) สำหรับโปรแกรม IBM SPSS ไม่สามารถหาค่า ความคลาดเคลื่อน ในการทำ backcasting ได้ในส่วนที่ทำนายไปข้างหน้าต้องคำนวณจากสูตร

ผลการพยากรณ์จำนวนผู้ป่วยโรคอาหารเป็นพิษ (food poisoning) จังหวัดกรุงเทพมหานคร ด้วยโปรแกรม IBM SPSS หาตัวแบบที่เหมาะสมที่สุดด้วยวิธีบ็อกซ์เจนกินส์ คือ ARIMA (1, 1, 1) × (1, 0, 1)¹² ค่าสถิติ Ljung-Box Q test มีค่า p value > 0.05 ไม่มีนัยสำคัญทางสถิติ และได้พยากรณ์ไปข้างหน้าตั้งแต่เดือนมกราคมถึงเดือนธันวาคม พ.ศ. 2562 ผลจากการเปรียบเทียบค่าข้อมูลจริงกับค่าพยากรณ์ด้วย backcasting ได้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 56.94, ค่าเปอร์เซ็นต์ผิดพลาดเฉลี่ยสัมบูรณ์ (MAPE) เท่ากับ 18.18 นำค่าประมาณของพารามิเตอร์ที่ได้จากตาราง Output ภาพที่ 4F มาแทนค่าลงในตัวแบบ ARIMA (1, 1, 1) × (1, 0, 1)¹² จะได้เป็นสมการการพยากรณ์ที่สามารถใช้ในการพยากรณ์ค่าในอนาคตได้ต่อไป เมื่อได้ค่าพยากรณ์ล่วงหน้าแล้ว สามารถใช้สร้างแนวทางการป้องกันการเกิดโรคดังกล่าวได้ ด้วยกรมควบคุมโรค ขอให้ประชาชนใช้มาตรการ “สุก ร้อน สะอาด” โดยรับประทานอาหารปรุงสุกใหม่ด้วยความร้อน ไม่มีแมลงวันตอม ล้างมือด้วยน้ำและสบู่บ่อยๆ รักษาสุขอนามัยจะช่วยป้องกันการโรคอาหารเป็นพิษได้

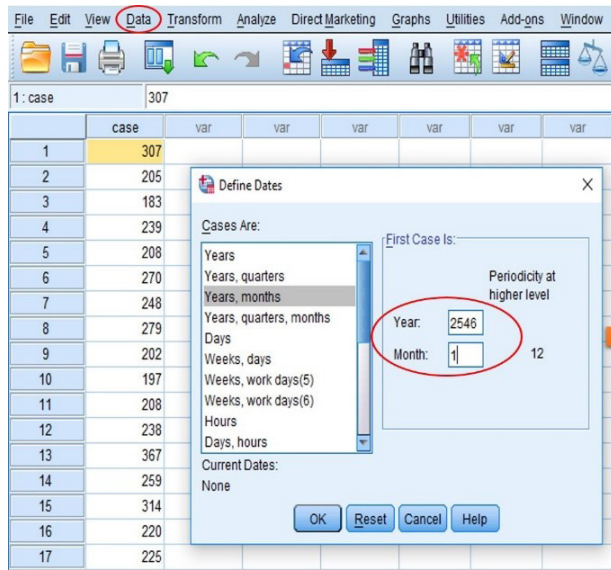
การวิเคราะห์ข้อมูลอนุกรมเวลา อาจมีตัวแบบที่เหมาะสมมากกว่า

1 ตัวแบบ ควรเลือกตัวแบบ เช่น ตัวแบบที่มีจำนวนพารามิเตอร์น้อยกว่าหรือตัวแบบที่สามารถนำมาใช้ในการพยากรณ์ได้ง่ายกว่า และมีค่าความคลาดเคลื่อน (error) น้อยกว่าด้วย

ตารางที่ 2 แสดงการเปรียบเทียบการตรวจสอบตัวแบบจากโปรแกรม IBM SPSS

SARIMA model	R square	RMSE	MAE	MAPE	Ljung Box Q	ผลการพิจารณา
ARIMA (1, 1, 1) × (1, 0, 1) ¹²	0.582	80.63	56.94	18.18	0.145	เป็น model ที่ดีที่สุด
ARIMA (1, 1, 1) × (1, 1, 1) ¹²	0.573	82.11	59.49	19.05	0.121	Error สูงกว่า
ARIMA (1, 1, 0) × (1, 1, 0) ¹²	0.314	103.42	73.91	22.83	0.000	Error สูงกว่า
ARIMA (1, 1, 1) × (1, 0, 0) ¹²	0.510	87.09	61.34	19.77	0.333	Error สูงกว่า
Expert modeler จากโปรแกรมคำนวณ	0.599	77.95	56.60	17.56	0.000	ค่า Ljung-Box test p value < 0.05

ที่มา: วรียา วงศ์พานิช

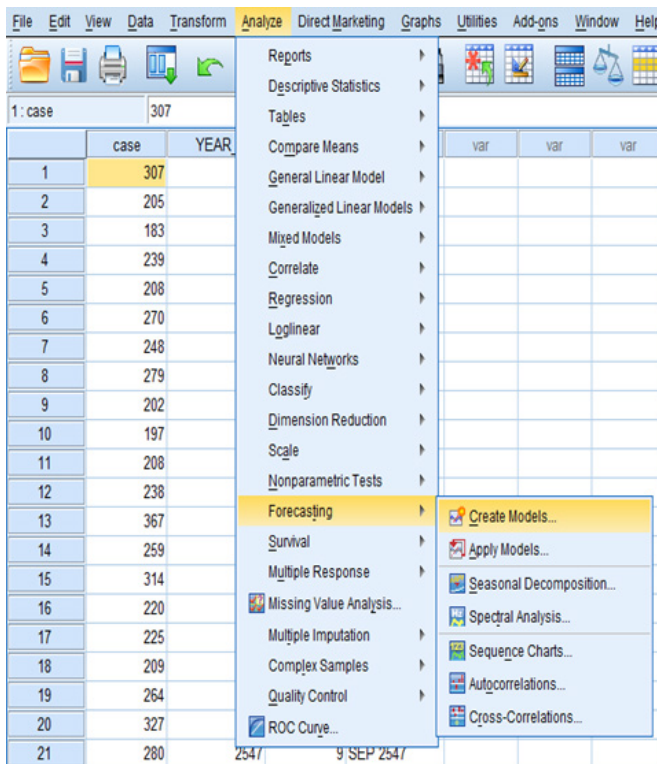


1A

	case	YEAR_	MONTH_	DATE_	var
1	307	2546	1	JAN 2546	
2	205	2546	2	FEB 2546	
3	183	2546	3	MAR 2546	
4	239	2546	4	APR 2546	
5	208	2546	5	MAY 2546	
6	270	2546	6	JUN 2546	
7	248	2546	7	JUL 2546	
8	279	2546	8	AUG 2546	
9	202	2546	9	SEP 2546	
10	197	2546	10	OCT 2546	
11	208	2546	11	NOV 2546	
12	238	2546	12	DEC 2546	
13	367	2547	1	JAN 2547	
14	259	2547	2	FEB 2547	
15	314	2547	3	MAR 2547	
16	220	2547	4	APR 2547	
17	225	2547	5	MAY 2547	

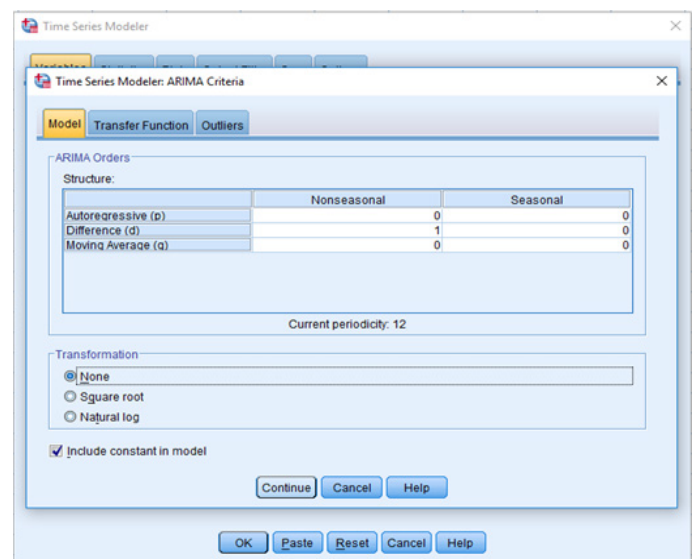
1B

ภาพที่ 1 การเตรียมข้อมูล
ที่มา: วรียา วงศ์พานิช

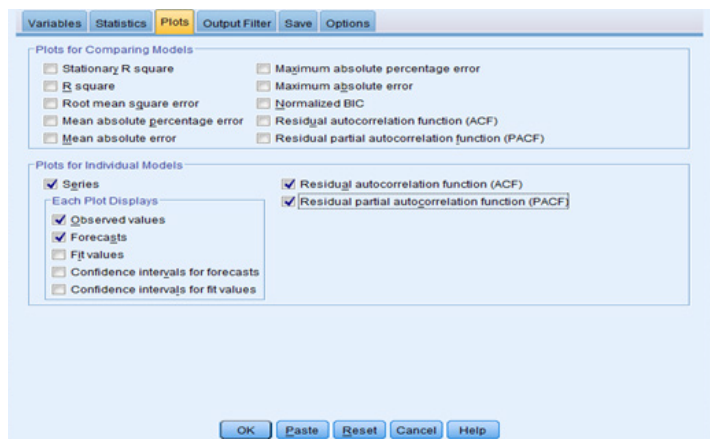


2A

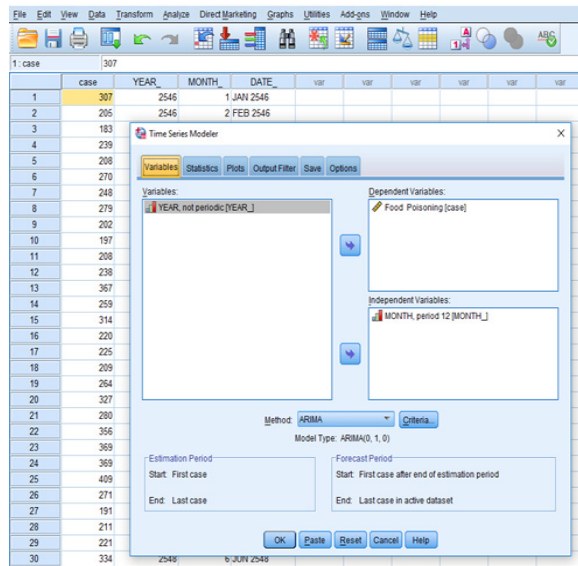
ภาพที่ 2 แสดงการแปลงข้อมูลอนุกรมเวลา
ที่มา: วรียา วงศ์พานิช



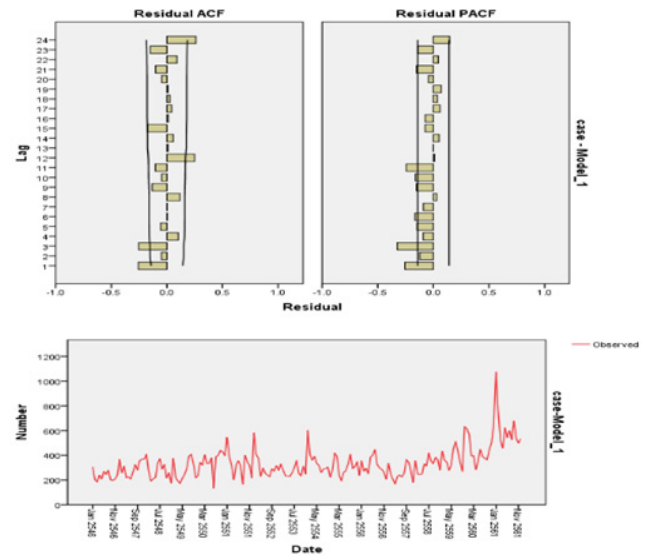
2B



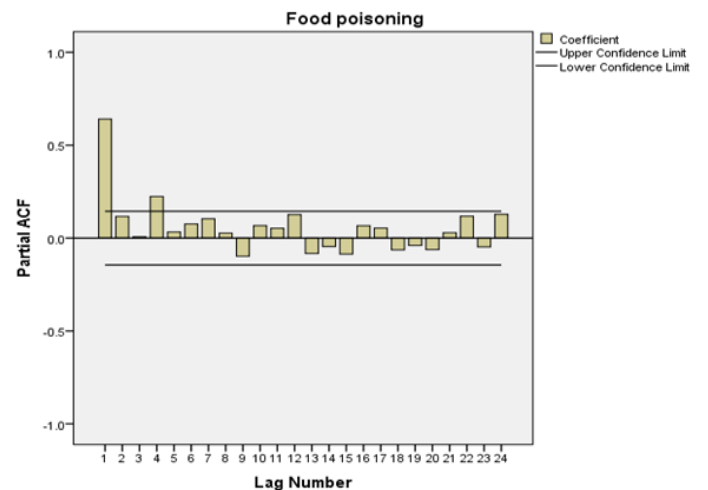
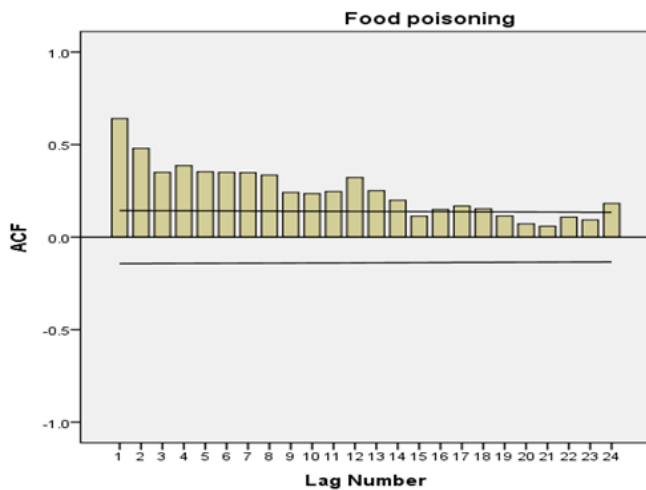
2C



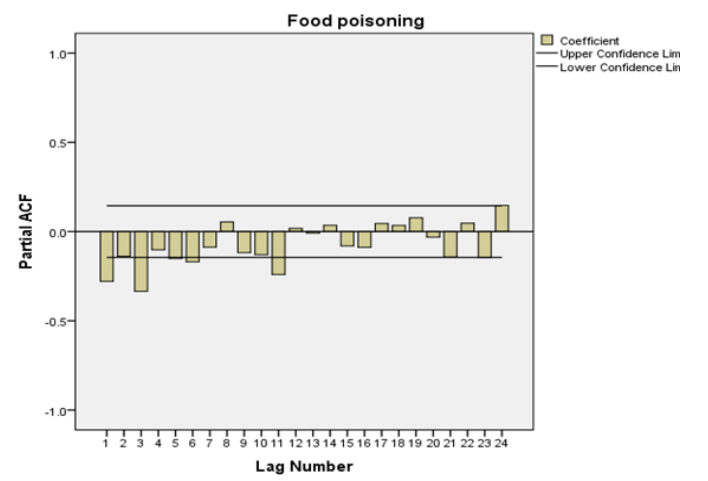
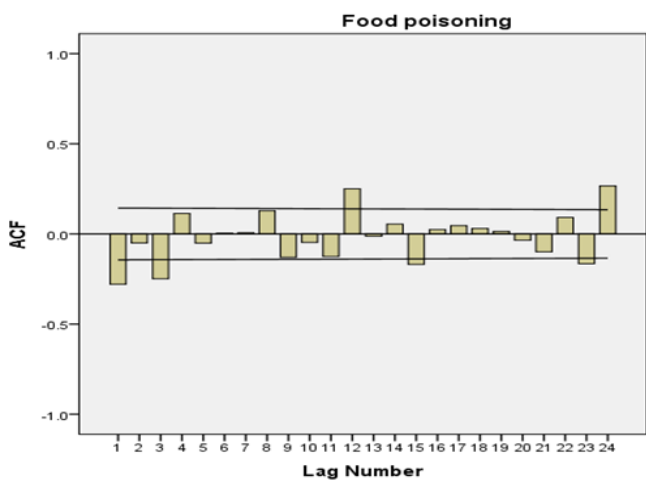
2D



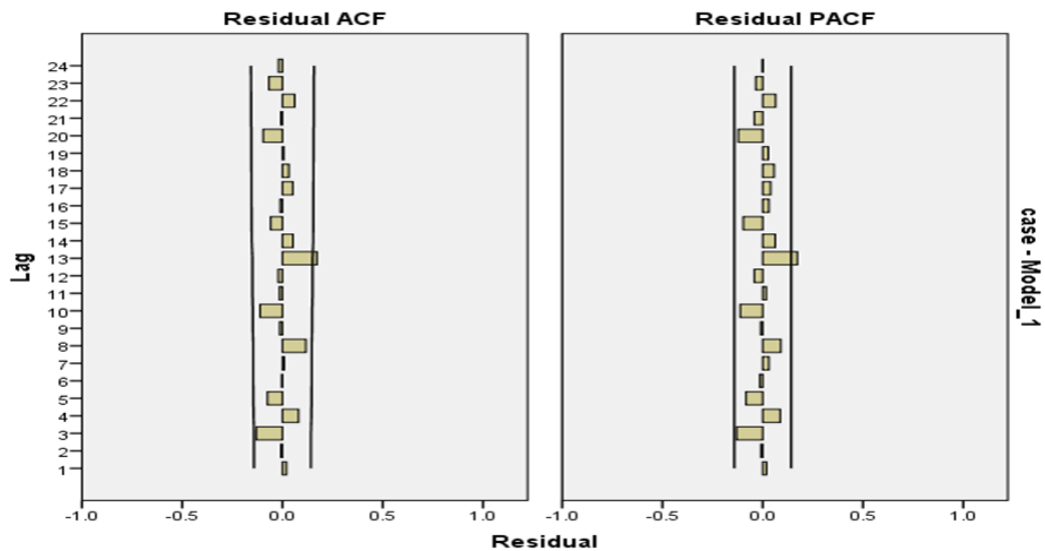
2E



ภาพที่ 3A กราฟ ACF และ PACF ของอนุกรมเวลาของการป่วยด้วยโรคอาหารเป็นพิษ
ที่มา: วรียา วงศ์พานิช



ภาพที่ 3B กราฟ ACF และ PACF ของอนุกรมเวลาของการป่วยด้วยโรคอาหารเป็นพิษ
เมื่อแปลงข้อมูลด้วยการหาผลต่างแนวโน้มลำดับที่ 1
ที่มา: วรียา วงศ์พานิช



ภาพที่ 3C กราฟ ACF และ PACF ของความคลาดเคลื่อนจากการพยากรณ์โดยวิธีบ็อกซ์เจนกินส์ ที่มีตัวแบบ SARIMA(1,1,1)(1,0,1)¹²
ที่มา: วรียา วงศ์พานิช

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help



13:

	case	YEAR
1	307	
2	205	
3	183	
4	239	
5	208	
6	270	
7	248	
8	279	
9	202	
10	197	
11	208	
12	238	
13	367	
14	259	
15	314	
16	220	
17	225	
18	209	
19	264	
20	327	
21	280	2547
22	356	2547
23	369	2547
24	369	2547
25	409	2548
26	271	2548
27	191	2548
28	211	2548
29	221	2548
30	334	2548
31	373	2548
32	287	2548
33	328	2548
34	217	2548
35	261	2548
36	175	2548
37	376	2549

Reports

Descriptive Statistics

Tables

Compare Means

General Linear Model

Generalized Linear Models

Mixed Models

Correlate

Regression

Loglinear

Neural Networks

Classify

Dimension Reduction

Scale

Nonparametric Tests

Forecasting

Survival

Multiple Response

 Missing Value Analysis...

Multiple Imputation

Complex Samples

Quality Control

 ROC Curve...

 Create Models...

 Apply Models...

 Seasonal Decomposition...

 Spectral Analysis...

 Sequence Charts...

 Autocorrelations...

 Cross-Correlations...

var

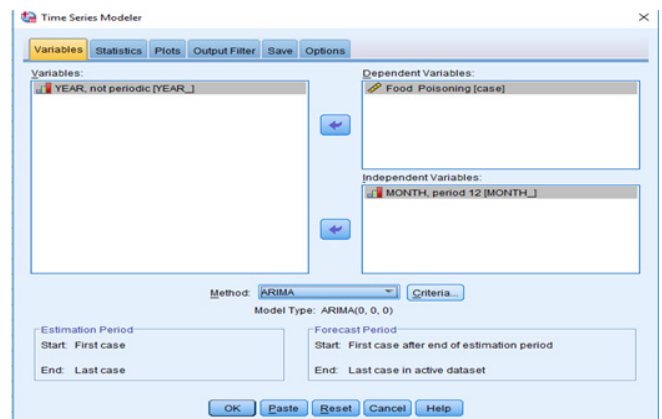
var

var

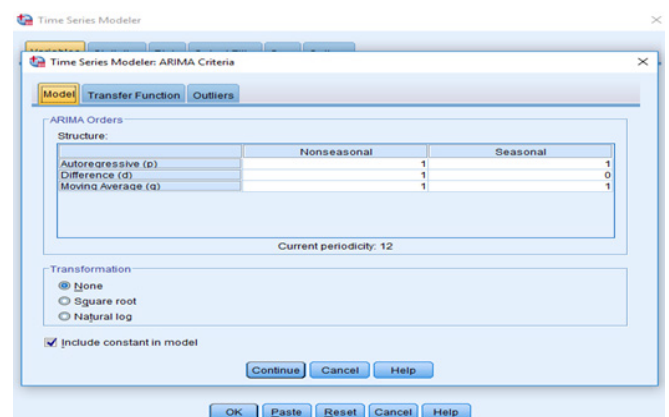
Data View

Variable View

4A

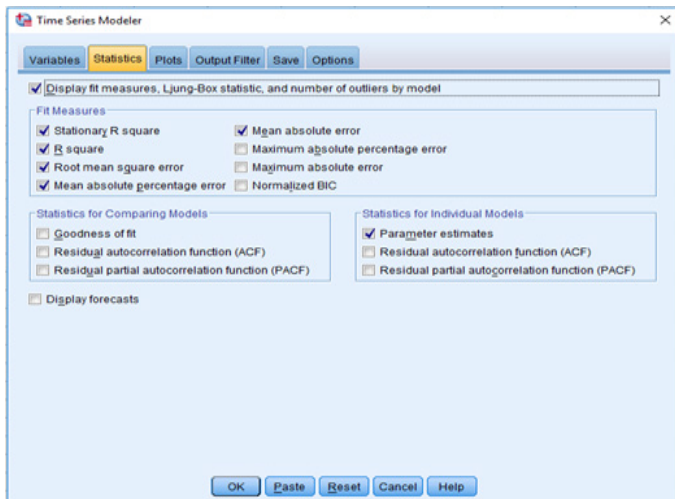


4B

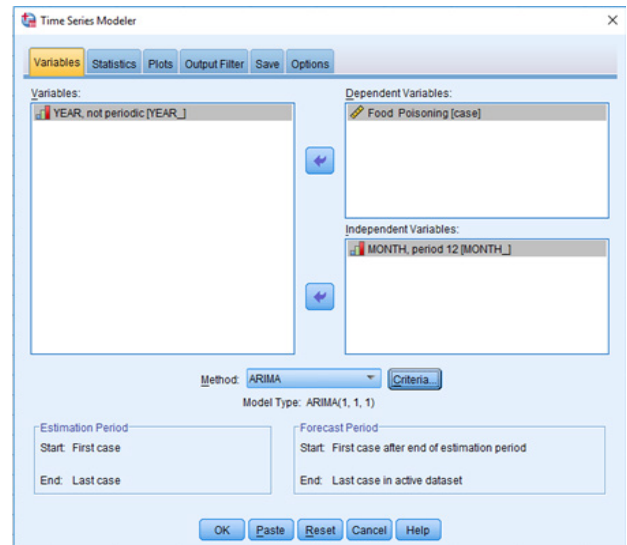


4C

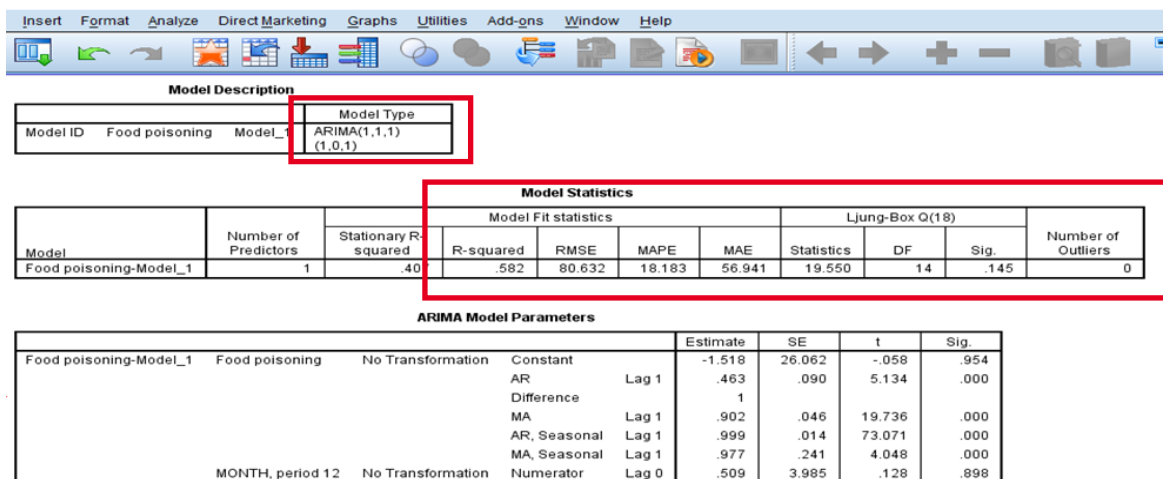
ภาพที่ 4 แสดงการทดสอบ model ที่เลือกมาได้
ที่มา: วรียา วงศ์พานิช



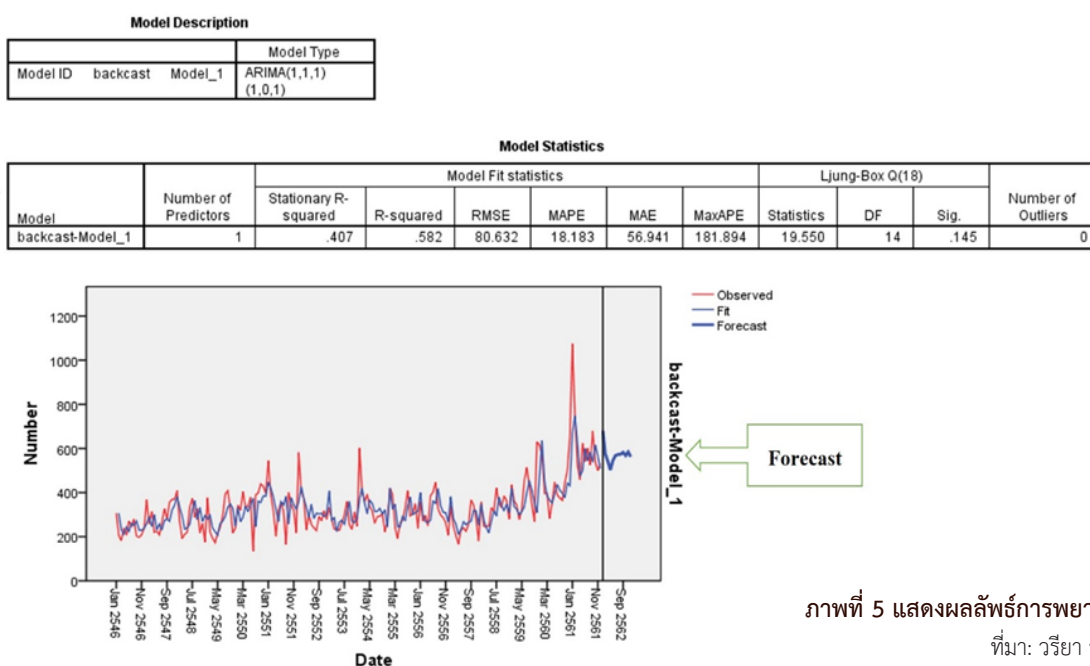
4D



4E



4F



ภาพที่ 5 แสดงผลลัพธ์การพยากรณ์ย้อนหลัง backcasting
ที่มา: วรวิภา วงศ์พานิช

SPSS มีคำสั่ง Expert Modeler ที่ช่วยให้เราเลือกตัวแบบได้สะดวกขึ้น โดยโปรแกรมจะเลือกตัวแบบที่เหมาะสมกับข้อมูลที่สุ่มมาให้อย่างไรก็ตามผู้พยากรณ์ควรนำตัวแบบที่ได้มาทดสอบเปรียบเทียบกับตัวแบบอื่นๆ (ที่มีค่า fit statistics ใกล้เคียงกัน) โดยการทำการ backcasting ตรวจสอบความเหมาะสมในการพยากรณ์ด้วย เนื่องจากโปรแกรมจะให้ตัวแบบที่เหมาะสมที่สุดอาจจะไม่ใช่ตัวแบบที่พยากรณ์ไปข้างหน้าได้ดีที่สุด ควรเลือกตัวแบบด้วยตัวเอง ไม่ควรใช้คำสั่ง Expert Modeler โดยไม่ตรวจสอบค่าสถิติเพราะจะได้ข้อมูลที่ผิดเพี้ยน มีค่า error เยอะมาก สำหรับข้อมูลชุดนี้ Expert Modeler ให้ค่า Ljung-Box Q test p value < 0.05 ถึงแม้ว่าค่าความคลาดเคลื่อน (error) จะน้อยกว่าก็ตาม

สรุป

การพยากรณ์ข้อมูลอนุกรมเวลาทางการแพทย์ด้วยวิธีบ็อกซ์-เจนกินส์ตัวแบบอาร์มีมีเงื่อนไขว่าจะต้องใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลาที่มีคุณสมบัติคงที่เท่านั้น และมีฤดูกาลได้แต่ต้องมีคุณสมบัติคงที่เช่นเดียวกัน ในกรณีที่ข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์มีคุณสมบัติไม่คงที่ จะต้องทำการแปลงข้อมูลอนุกรมเวลา จากการพิจารณาว่าข้อมูลอนุกรมเวลานั้นว่ามีค่าเฉลี่ยไม่คงที่ (มีแนวโน้มเป็นองค์ประกอบ) หรือความแปรปรวนของอนุกรมเวลาไม่คงที่ วิธีการแปลงข้อมูลก็จะแตกต่างกัน โดยการหาผลต่างสำหรับข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ยไม่คงที่ และการใช้ฟังก์ชันการแปลงข้อมูลสำหรับความแปรปรวนไม่คงที่ก่อนที่จะนำข้อมูลไปใช้สร้างตัวแบบอาร์มี และในปัจจุบันได้มีการพัฒนาทางด้านโปรแกรมสำเร็จรูปสำหรับการพยากรณ์เชิงปริมาณหลากหลายโปรแกรมที่สร้างความสะดวก รวดเร็ว และถูกต้องแม่นยำ แสดงค่าวัดความคลาดเคลื่อนไว้เพื่อให้ผู้พยากรณ์ใช้ความรู้ในการตัดสินใจเลือกวิธีพยากรณ์ตามหลักเกณฑ์ที่กำหนดไว้ โดยเลือกวิธีการพยากรณ์ที่เหมาะสมกับลักษณะข้อมูลให้ค่าความคลาดเคลื่อนสำหรับการพยากรณ์ต่ำที่สุด การพยากรณ์ด้วยวิธีบ็อกซ์-เจนกินส์ตัวแบบอาร์มี มี 4 ขั้นตอนโดยใช้โปรแกรม IBM SPSS เป็นทางเลือกหนึ่งที่ทำให้ความแม่นยำสูงสำหรับผู้พยากรณ์ที่ขาดประสบการณ์สามารถนำมาพยากรณ์สำหรับข้อมูลอนุกรมเวลาที่มีความผันแปรของ

แนวโน้ม วัฏจักร และฤดูกาลที่มีความไม่ชัดเจน สร้างตัวแบบในการพยากรณ์ที่เหมาะสมได้

การรายงานผลการวิเคราะห์ข้อมูลอนุกรมเวลานอกจากการระบุตัวแบบและค่าความคลาดเคลื่อนต่างๆแล้ว สำหรับการเขียนรายงานฉบับสมบูรณ์ให้แสดงสมการการพยากรณ์ของตัวแบบ ซึ่งประกอบด้วยค่า Coefficient ขององค์ประกอบต่างๆ ได้แก่ AR, MA, SAR, SMA แต่ละตัว และนัยสำคัญทางสถิติด้วย ซึ่งค่าต่างๆ เหล่านี้จะแสดงอยู่ตาราง ARIMA Model Parameters ของ IBM SPSS Output สมการการพยากรณ์ที่ได้มาสำหรับข้อมูลชุดนั้นๆ สามารถนำมาใช้พยากรณ์ข้อมูลอนุกรมเวลาในอนาคตได้ต่อไป

กิตติกรรมประกาศ

ขอขอบคุณสำนักโรคระบาด กรมควบคุมโรคที่ให้ความอนุเคราะห์ข้อมูลอนุกรมเวลาจำนวนผู้ป่วยรายเดือนโรคอาหารเป็นพิษ (food poisoning) ในการนำมาเป็นตัวอย่างการวิเคราะห์ข้อมูลสถิติพยากรณ์วิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบอาร์มีสำหรับบทความนี้

เอกสารอ้างอิง

1. สำนักโรคระบาด. จำนวนผู้ป่วย-เสียชีวิตรายเดือนโรคอาหารเป็นพิษ (Food poisoning) [อินเทอร์เน็ต]. 2563. เข้าถึงเมื่อ 1 กรกฎาคม 2563]. เข้าถึงได้จาก <http://www.boe.moph.go.th/boedb/surdata/>
2. ทรงศิริ แด่สมบัติ. เทคนิคการพยากรณ์เชิงปริมาณ. ภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพมหานคร: สำนักพิมพ์ฟิสิกส์เซ็นเตอร์, 2539.
3. Box, G. E., G. M. Jenkins, and G. C. Reinsel. 1994. Time series analysis: Forecasting and control. (3rd ed) Englewood Cliffs, NJ: Prentice-Hall. New Jersey.
4. Makridakis, S., Wheelwright, S. C. and McGee, V.E. Forecasting Method and Application 2nd edition, New York: John Wiley & Son; 1983.
5. สุพรรณิ อึ้งปัญสดวงค์. เทคนิคการพยากรณ์เชิงสถิติ. ขอนแก่น: โรงพิมพ์มหาวิทยาลัยขอนแก่น, 2555.
6. ศิริลักษณ์ สุวรรณวงศ์. เทคนิคการพยากรณ์เชิงปริมาณ: การวิเคราะห์อนุกรมเวลา. นครปฐม: สำนักพิมพ์มหาวิทยาลัยมหิดล, 2556.
7. Vongspanich, J. A Comparison on Statistical Techniques between Single and Combined Forecast: A Case Study on Traffic Accident Data in Bangkok. [M.S.Thesis in Biostatistics]. Bangkok: Faculty of Graduate Studies, Mahidol University, 2000.