นิพนธ์ต้นฉบับ

ตัวแบบการพยากรณ์อุบัติการณ์การเกิดโรคมาลาเรียในจังหวัดอุบลราชธานี จากข้อมูลผู้ป่วยรายเดือนและปัจจัยด้านสภาวะอากาศ

กิตติ์ธเนศ นิธิวรเสฏฐ์[1], จิราพร เขียวอยู่[2] และสุพรรณี อึ้งปัญสัตวงศ์[3]

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อหาตัวแบบการพยากรณ์อุบัติการณ์การเกิดโรคมาลาเรียที่เหมาะสมที่สุดของพื้นที่จังหวัดอุบลราชธานี โดยเปรียบเทียบระหว่างตัวแบบการพยากรณ์เชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยปัวส์ซอง ที่มีตัวแปรตามเป็นจำนวนผู้ป่วยโรคมาลาเรีย ตัวแปรทำนายเป็นตัวแปรสภาวะอากาศ และตัวแบบการพยากรณ์ของบอกซ์และเจนกินส์ จากข้อมูลอัตราป่วยต่อประชากรแสนคน ขั้นตอนการพัฒนาตัวแบบใช้ข้อมูลขนาด 108 ค่า ขั้นตอนการประเมินตัวแบบใช้ข้อมูลขนาด 12 ค่า

ผลการศึกษาพบว่า การวิเคราะห์ด้วยตัวแบบการพยากรณ์เชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยปัวส์ซอง เกิดปัญหาตัวแปรตามมีความแปรปรวนมากกว่าค่าเฉลี่ย จึงแก้ไขโดยใช้ตัวแบบเชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยทวินามเชิงลบ พบว่า ตัวแบบที่ดีที่สุด มีเวลา และอุณหภูมิสูงสุด เป็นตัวแปรทำนาย ในขั้นตอนการพัฒนาตัวแบบ มีค่า MAD, MSE, และ MAPE เท่ากับ 12.33, 288.53, และ 24.56 ตามลำดับ ในขั้นตอนตรวจสอบความถูกต้องของการพยากรณ์ มีค่า MAD, MSE, และ MAPE เท่ากับ 19.04, 558.22, และ 33.46 ตามลำดับ ตัวแบบการพยากรณ์ของบอกซ์และเจนกินส์ที่ได้ คือ ARIMA(1,0,0) ในขั้นตอนการพัฒนาตัวแบบ มีค่า MAD, MSE, และ MAPE เท่ากับ 23.28, 1179.30, และ 38.33 ตามลำดับ ในขั้นตอนตรวจสอบความถูกต้องของการพยากรณ์ มีค่า MAD, MSE, และ MAPE เท่ากับ 19.45, 590.09, และ 53.00 ตามลำดับ

ซึ่งสรุปได้ว่าตัวแบบเชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยทวินามเชิงลบเป็นตัวแบบที่เหมาะสมที่สุด เพราะให้ค่าความคลาดเคลื่อนต่ำที่สุดไม่ว่าจะพิจารณาจากเกณฑ์ใด และตัวแบบนี้น่าจะมีความเหมาะสมในการนำไปใช้ในทางปฏิบัติ เพราะอาศัยตัวแปรด้านสภาวะอากาศในการทำนายเพียงตัวแปรเดียว คือ อุณหภูมิสูงสุด และเป็นข้อมูลที่หาได้ง่าย

คำสำคัญ: *การพยากรณ์, อุบัติการณ์การเกิดโรคมาลาเรีย, สมการถดถอยทวินามเชิงลบ*

[1] *นักศึกษาหลักสูตรสาธารณสุขศาสตรบัณฑิต สาขาวิชาชีวสถิติ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น* (e-mail: oodkit@hotmail.com)
[2] *ผู้ช่วยศาสตราจารย์ ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น*
[3] *รองศาสตราจารย์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น*

Original Article

Modeling Malaria Incidence in Ubonratchatani Province Using Monthly Case Reports and Weather Conditions

Kittanate Nithiworaset [(1)], *Jiraporn Khiewyoo* [(2)] *and Supunnee Ungpansattawong* [(3)]

Abstract

This study aimed to determine the best model for forecasting malaria incidence in the Ubonratchatani province. A comparison between the generalized linear models (GLM) Poisson regression and the Box-Jenkins techniques was considered. For the Poisson regression, the number of incidences of malaria cases was taken as the response variable and weather records were used as predictors. The incidence rate per 100,000 population was used for the Box and Jenkins technique. In developing the model, a data size of 108 was used, whereas a data size of 12 was used for the validation.

GLM Negative binomial regression was superior to the GLM Poisson regression because of an overdispersion problem. Time and maximum temperature were predictors in the GLM Negative binomial regression. In the model development, the MAD, MSE, and MAPE of the GLM Negative binomial regression were 12.33, 288.53, and 24.46 respectively. In the validation, MAD, MSE, and MAPE were 19.04, 558.22, and 33.46 respectively. The best model from the Box and Jenkins technique was ARIMA (1,0,0). In the model development, MAD, MSE, and MAPE of ARIMA (1,0,0) were 23.28, 1179.30, and 38.33 respectively. For the validating step MAD, MSE, and MAPE of ARIMA (1,0,0) were 19.45, 590.09, and 53.00 respectively.

In conclusion, GLM Negative binomial regression was the best forecasting model of malaria incidence for the Ubonratchatani province because of the lowest error and because it needs only one weather predictor, maximum temperature, which can be easily obtained.

Keyword: *forecasting, malaria incidence, negative binomial regression model*

[(1)] *Master Degree in Public Health Student, Faculty of Public Health, Khon Kaen University* (e-mail: oodkit@hotmail.com)
[(2)] *Assistant Professor, Department of Biostatistics and Demography, Faculty of Public Health, Khon Kaen University*
[(3)] *Associate Professor, Department of Statistics, Faculty of Science, Khon Kaen University*

บทนำ

มาลาเรียเป็นโรคติดต่อนำโดยแมลง พบการระบาดครั้งแรกเมื่อปี ค.ศ.1940 ในประเทศไทยปัจจุบัน พบว่ามีจำนวนผู้ป่วยมาลาเรียทั้งประเทศในเดือนมกราคม 2552 – ธันวาคม 2552 จำนวนทั้งสิ้น 23,345 ราย คิดเป็นอัตราป่วย 0.32 ต่อประชากรหนึ่งพันคน (สำนักโรคติดต่อนำโดยแมลง, 2552) จังหวัดอุบลราชธานี พบผู้ป่วยในเดือนมกราคม 2552 ถึงธันวาคม 2552 จำนวนทั้งสิ้น 626 ราย คิดเป็น 0.35 ต่อประชากรหนึ่งพันคน ส่วนใหญ่อยู่ในเขตอำเภอน้ำยืน นาจะหลวย บุญฑริก น้ำขุ่น โพธิ์ไทรและสิรินธร (สำนักงานควบคุมและป้องกันโรคที่ 7, 2552) ผลกระทบของโรคไข้มาลาเรีย องค์การอนามัยโลกรายงานว่า ทุกปีประชากรโลก 400 ล้านคน จะล้มป่วยเป็นไข้และมีอาการหนาวสั่นด้วยโรคมาลาเรียและคน 2 ล้านคนจะเสียชีวิตด้วยโรคนี้ โดยเหยื่อของโรคนี้ส่วนมากมักเป็นเด็กที่มีอายุน้อยกว่า 5 ขวบ ในกรณีของครอบครัวที่ยากจน หากหัวหน้าครอบครัวล้มป่วยด้วยโรคมาลาเรีย จะทำให้ขาดรายได้ที่จะนำมาจุนเจือครอบครัว หากเกิดการระบาดที่ประเทศใด การท่องเที่ยว การค้า ซึ่งเป็นแหล่งที่มาของเงินรายได้จากนักทัศนาจรจะลดลง ทำให้เศรษฐกิจตกต่ำลง (บุญส่ง สอนส่งเสริม, 2553) วิธีการป้องกันและควบคุมโรค ผู้ที่เกี่ยวข้องต้องคำนึงถึงปัจจัยที่ก่อให้เกิดโรคคือ คน ยุง เชื้อและสิ่งแวดล้อม การประมาณจำนวนผู้ป่วยล่วงหน้าที่แน่นอน จะทำให้สามารถวางแผนในการป้องกันและควบคุมไม่ให้คนเกิดการป่วยด้วยโรคมาลาเรียได้ดีขึ้น ดังนั้นการใช้การพยากรณ์ ซึ่งเป็นวิธีการวิเคราะห์ข้อมูลอนุกรมเวลาจึงมีบทบาทและมีการนำมาใช้มากขึ้นในงานวิจัยด้านระบาดวิทยาในปัจจุบัน (Gaudart et al., 2009) รวมถึงการมีบทบาทในการกำหนดยุทธศาสตร์ ในการวางแผนป้องกันโรคต่าง ๆ มากขึ้น ตัวแบบพยากรณ์ที่ให้ค่าพยากรณ์ที่ถูกต้อง แม่นยำมากที่สุด จะเป็นตัวแบบที่ถูกนำไปใช้ในการพยากรณ์ เพื่อวางแผนป้องกันโรคได้อย่างมีประสิทธิภาพ ทำให้หน่วยงานด้านสาธารณสุข ใช้งบประมาณได้อย่างคุ้มค่าไม่สิ้นเปลือง ผู้บริหารได้รับข้อมูลสารสนเทศที่เชื่อถือได้สามารถกำหนดนโยบายวางแผนได้อย่างถูกต้อง แต่ในประเทศไทย การประยุกต์ใช้การพยากรณ์ในด้านสาธารณสุขและระบาดวิทยายังไม่มีการนำมาใช้อย่างแพร่หลาย

วัตถุประสงค์การวิจัย

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อหาตัวแบบที่เหมาะสมที่สุดในการพยากรณ์อุบัติการณ์การเกิดโรคมาลาเรียในจังหวัดอุบลราชธานี โดยเปรียบเทียบระหว่างตัวแบบการพยากรณ์ แบบเชิงเส้นโดยนัยทั่วไป (Generalized Linear Models: GLMs) ประเภทตัวแบบถดถอยปัวส์ซอง (Poisson Regression) ซึ่งใช้ตัวแปรด้านสภาวะอากาศหลาย ๆ ตัวแปรในการทำนายกับตัวแบบการพยากรณ์ของบอกซ์และเจนกินส์ (Box and Jenkins model) ซึ่งเป็นตัวแบบการพยากรณ์ที่อาศัยข้อมูลอุบัติการณ์การเกิดโรคมาลาเรียจากอดีตเพียงอย่างเดียว

วิธีดำเนินการวิจัย

พื้นที่การศึกษา

พื้นที่ในการศึกษาครั้งนี้ คือ จังหวัดอุบลราชธานี ซึ่งตั้งอยู่ทางตะวันออกเฉียงเหนือของประเทศ อยู่ห่างจากกรุงเทพมหานครประมาณ 630 กิโลเมตร มีเนื้อที่ประมาณ 15,739 ตารางกิโลเมตร มีประชากรทั้งสิ้น 1,803,754 คน

การประมวลผลข้อมูล

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ คือ

1) ข้อมูลผู้ป่วยโรคไข้มาลาเรีย เป็นข้อมูลทุติยภูมิได้จากการดึงจากฐานข้อมูลวิเคราะห์สถานการณ์โรครายเดือนย้อนหลัง ระหว่างเดือน มกราคม 2543 ถึง เดือนธันวาคม 2552 จากสำนักโรคติดต่อนำโดยแมลง กรมควบคุมโรค กระทรวงสาธารณสุข โดยแบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 เป็นข้อมูลสำหรับขั้นตอนการพัฒนาตัวแบบระหว่างเดือนมกราคม 2543 ถึง เดือนธันวาคม 2551 จำนวน 108 เดือน และชุดที่ 2 เป็นข้อมูลสำหรับขั้นตอนการตรวจสอบตัวแบบระหว่างเดือนมกราคม 2552 ถึงเดือนธันวาคม 2552 จำนวน 12 เดือน (สำนักโรคติดต่อนำโดยแมลง, 2552)

2) ข้อมูลสภาวะอากาศ ข้อมูลสภาวะอากาศที่นำมาศึกษา ประกอบด้วยปริมาณน้ำฝน อุณหภูมิและความชื้นสัมพัทธ์จากศูนย์อุตุนิยมวิทยาภาคตะวันออกเฉียงเหนือจังหวัดอุบลราชธานีเป็นข้อมูลรายเดือนขนาด 120 เดือน โดยแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดที่ 1 เป็นข้อมูลสำหรับขั้นตอนการพัฒนาตัวแบบระหว่างเดือนมกราคม 2543 ถึงเดือนธันวาคม 2551 จำนวน 108 เดือน และ ชุดที่ 2 เป็นข้อมูลสำหรับขั้นตอนการตรวจสอบตัวแบบระหว่างเดือนมกราคม 2552 ถึงเดือนธันวาคม 2552 จำนวน 12 เดือน

(ศูนย์อุตุนิยมวิทยาภาคตะวันออกเฉียงเหนือ, 2552)

3) การตรวจสอบความถูกต้องของข้อมูล ดำเนินการโดยการนำข้อมูลที่ได้บันทึกไว้ในไฟล์อิเล็คทรอนิคเปรียบเทียบกับแบบบันทึกข้อมูลเพื่อความตรงกันของข้อมูล ผลการตรวจสอบพบว่าข้อมูลมีความตรงกันและถูกต้องตลอดจนไม่พบค่าผิดปกติของข้อมูล

การวิเคราะห์ข้อมูล

การศึกษาครั้งนี้ จำแนกการวิเคราะห์ข้อมูลหรือการสร้างตัวแบบพยากรณ์อุบัติการณ์การเกิดโรคมาลาเรียตามประเภทของข้อมูล เป็น 2 วิธี คือ

วิธีที่ 1 การวิเคราะห์ตัวแบบพยากรณ์ วิธีเชิงเส้นโดยนัยทั่วไป ประเภทตัวแบบถดถอยปัวส์ซอง (Generalized Linear Models: Poisson Regression) ที่มีตัวแปรตามเป็นข้อมูลจำนวนนับ (Count Data) และมีตัวแปรทำนายเป็นตัวแปรสภาวะอากาศ 3 ตัวแปร คือ ตัวแปรอุณหภูมิสูงสุด อุณหภูมิต่ำสุดและปริมาณน้ำฝน ซึ่งตัวแปรทั้งสามนี้เป็นตัวแปรที่คาดว่าจะมีความสัมพันธ์กับการเกิดโรคมาลาเรีย และได้รวมตัวแปรด้านเวลาเป็นตัวทำนายอีก 2 ตัวแปร คือ เวลาและฤดูกาล ประเมินตัวแบบโดย

1) พิจารณาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามทีละตัว และโดยรวม จากสถิติ Wald test และ Log Likelihood Ratio test

2) ตรวจสอบความเหมาะสมของรูปแบบ (Goodness of fit) จากค่า Deviance และ Akaike information criterion (AIC)

3) วิเคราะห์ความคลาดเคลื่อน (Residual Analysis) ด้วยสถิติเพียร์สันไคสแควร์ “Pearson Chi-square statistic” (Hardin & Hilbe, 2001)

4) ตรวจสอบการกระจายของตัวแปรตาม หรือการเกิดโรคมาลาเรียว่า มีการกระจายมากเกินไปหรือไม่ (ตัวแปรตามมีความแปรปรวนมากกว่าค่าเฉลี่ย: Overdispersion) จากค่าความคลาดเคลื่อนของเพียร์สัน (Pearson Residual) กับองศาอิสระของค่าความคลาดเคลื่อน (Residual degree of freedom) และจากสถิติสกอร์ (score test) (Hardin & Hilbe, 2001)

วิธีที่ 2 การวิเคราะห์เพื่อหาตัวแบบการพยากรณ์โดยเทคนิคของบอกซ์และเจนกินส์ (Box-Jenkins) ตัวแปรที่นำมาวิเคราะห์ คือ อัตราป่วยต่อแสนประชากรรายเดือน ซึ่งเป็นข้อมูลต่อเนื่อง (Continuous Data) การสร้างตัวแบบโดยเทคนิคของ Box-Jenkins หรือตัวแบบอารีม่า (ARIMA Model : Autoregressive Integrated Moving Average Model) ซึ่งใช้ข้อมูลเพียงตัวแปรเดียว (ข้อมูลอนุกรมชุดเดียว) ไม่มีตัวแปรอิสระ โดยการหาตัวแบบอารีม่า อันดับ p,d,q ที่เหมาะสมนั้น พิจารณาจากฟังก์ชันสหสัมพันธ์ในตนเอง (Autocorrelation Function :ACF) และฟังก์ชันสหสัมพันธ์ในตนเองบางส่วน (Partial Autocorrelation Function :PACF) ของข้อมูลอนุกรมเวลาที่อยู่ในสภาวะคงที่ (Stationary) อันหมายถึง อนุกรมเวลาที่มีทั้งค่าเฉลี่ยและค่าความแปรปรวนที่คงที่ หากข้อมูลที่นำมาวิเคราะห์ไม่อยู่ในสภาวะคงที่ ต้องแปลงข้อมูลเพื่อปรับให้อนุกรมเวลานั้นอยู่ในสภาวะคงที่ก่อนการวิเคราะห์

การตรวจสอบความเหมาะสมของตัวแบบ

1) ทดสอบความเป็นอิสระของความคลาดเคลื่อนโดยใช้สถิติทดสอบ Box และ Ljung

2) ทดสอบค่าเฉลี่ยของความคลาดเคลื่อนเท่ากับ 0 โดยใช้ สถิติทดสอบ t-test

3) ทดสอบการแจกแจงของความคลาดเคลื่อนว่ามีการแจกแจงเป็นแบบปกติหรือไม่ โดยใช้สถิติทดสอบ Shapiro-Wilk test

4) ทดสอบความคงที่ของความแปรปรวนของความคลาดเคลื่อน โดยการพิจารณาแผนภาพการกระจายของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

การพยากรณ์ จะแปลงค่าพยากรณ์จากอัตราป่วยต่อประชากรแสนคนมาเป็นจำนวนผู้ป่วยให้เป็นจำนวนนับเพื่อให้เป็นหน่วยเดียวกันกับการพยากรณ์ด้วยวิธีที่ 1

การเปรียบเทียบตัวแบบพยากรณ์ ใช้เกณฑ์ทางสถิติต่อไปนี้

(1) Mean absolute deviation (MAD) คำนวณจาก

$$MAD = \frac{\sum |e_t|}{n}$$

โดย e_t = ผลต่างระหว่างค่าสังเกตกับค่าพยากรณ์ ณ เวลา t

n = ขนาดของอนุกรมเวลา

(2) Mean square error (MSE) คำนวณจาก

$$MSE = \frac{\sum e_t^2}{n}$$

โดย e_t = ผลต่างระหว่างค่าสังเกตกับค่าพยากรณ์ ณ เวลา t

n = ขนาดของอนุกรมเวลา

(3) Mean absolute percent error (MAPE) คำนวณจาก

$$MAPE = \frac{\sum | e_t / Y_t |}{n}$$

โดย e_t = ผลต่างระหว่างค่าสังเกตกับค่าพยากรณ์ ณ เวลา t

Y_t = ค่าสังเกต ณ เวลา t

n = ขนาดของอนุกรมเวลา

ผลการศึกษา

การเกิดโรคมาลาเรีย

การเกิดโรคมาลาเรียในจังหวัดอุบลราชธานีรายปีระหว่างปี 2543 ถึง ปี 2552 พบว่า ปี 2543 มีจำนวนผู้ป่วยมากที่สุดคือ 1,828 คน คิดเป็นอัตราป่วย 122.8 ต่อแสนประชากร โดยเดือนสิงหาคม 2543 มีจำนวนผู้ป่วยมากที่สุดคือ 264 คน คิดเป็นอัตราป่วย 17.74 ต่อแสนประชากรน้อยที่สุด ในปี 2548 พบผู้ป่วยจำนวน 283 คน คิดเป็นอัตราป่วย 15.7 ต่อแสนประชากร โดยพบว่าน้อยที่สุดในเดือน พฤษภาคม 2548 จำนวน 5 คน คิดเป็นอัตราป่วย 0.28 ต่อแสนประชากร (ดังรูปภาพที่ 1)

ปริมาณน้ำฝนโดยรวมของจังหวัดอุบลราชธานีระหว่างปี 2543 ถึง ปี 2552 พบว่า มีปริมาณน้ำฝนโดยรวมเฉลี่ย 1,652.0 มิลลิเมตร (ส่วนเบี่ยงเบนมาตรฐาน 241.2 มิลลิเมตร) อุณหภูมิเฉลี่ยแต่ละปีเฉลี่ย 27.7 องศาเซลเซียส (ส่วนเบี่ยงเบนมาตรฐาน 0.3 องศาเซลเซียส) อุณหภูมิสูงสุดเฉลี่ย 36.2 องศาเซลเซียส (ส่วนเบี่ยงเบนมาตรฐาน 0.7 องศาเซลเซียส) อุณหภูมิต่ำสุด เฉลี่ย 19.2 องศาเซลเซียส (ส่วนเบี่ยงเบนมาตรฐาน 0.8 องศาเซลเซียส) ความชื้นสัมพัทธ์เฉลี่ย 72.2 มิลลิบาร์ (ส่วนเบี่ยงเบนมาตรฐาน 0.8 มิลลิบาร์)

ตัวแบบเชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยปัวส์ซองและสมการถดถอยทวินามเชิงลบ

จากการเลือกตัวแปรเข้าสู่ตัวแบบสุดท้าย ด้วยวิธีขจัดออกทีละตัวแปร (Backward elimination) พบว่า ตัวแปรที่สามารถพยากรณ์การเกิดโรคมาลาเรียมี 4 ตัวแปร โดยตัวแปรที่มีความสำคัญสูงสุด คือ เวลา รองลงมาคือ ฤดูกาล(เดือน) อุณหภูมิเฉลี่ยและความชื้นสัมพัทธ์ มีรูปแบบสมการถดถอยปัวส์ซอง คือ

$$\log(\mu_t) = -10.93 - 0.02X_1 - 0.15X_{2.2} - 0.37X_{2.3} - 0.52X_{2.4} - 0.51X_{2.5} - 0.45X_{2.6} - 0.55X_{2.7} - 0.74X_{2.8} - 1.10X_{2.9} - 0.88X_{2.10} - 0.48X_{2.11} - 0.16X_{2.12} - 0.02X_3 + 0.04X_4 + \log(N)$$

โดยที่

$\log(\mu_t)$ = อัตราการเกิดโรคมาลาเรีย ณ เวลา t

X_1 = เวลา เป็นตัวแปรต่อเนื่อง

$X_{2.2}$ ถึง $X_{2.12}$ = เดือนกุมภาพันธ์ ถึง เดือนธันวาคม (แทนด้วย 1 กับ 0)

X_3= อุณหภูมิเฉลี่ย เป็นตัวแปรต่อเนื่อง (: องศาเซลเซียส)

X_4= ความชื้นสัมพัทธ์ เป็นตัวแปรต่อเนื่อง (: มิลลิบาร์)

จากการประเมินตัวแบบการพยากรณ์ ด้วยสถิติ Wald test พบว่า ตัวแปรเวลา ตัวแปรฤดูกาลและตัวแปรความชื้นสัมพัทธ์ มีความสัมพันธ์กับการเกิดโรคมาลาเรียที่ระดับนัยสำคัญ 0.05 (p-value < 0.05) ตัวแปรอุณหภูมิเฉลี่ยไม่มีความสัมพันธ์กับการเกิดโรคมาลาเรียที่ระดับนัยสำคัญ 0.05 (p-value > 0.05) แต่เมื่อทดสอบความสัมพันธ์ระหว่างตัวแปรทำนายโดยรวมกับการเกิดโรคมาลาเรีย โดยสถิติ Likelihood Ratio test พบว่า โดยรวมแล้วตัวแบบสามารถอธิบายถึงความสัมพันธ์ระหว่างตัวแปรทำนายกับการเกิดโรคมาลาเรียได้ที่ระดับนัยสำคัญ 0.05 (LR chi-square,df=15 เท่ากับ 2021.59, p-value < 0.01) การตรวจสอบความเหมาะสมของรูปแบบ (Goodness of fit) ของตัวแบบพยากรณ์การเกิดโรคมาลาเรียจากค่า Deviance พบว่า ตัวแบบการทำนายไม่มีความเหมาะสมของรูปแบบที่ระดับนัยสำคัญ 0.05 (Goodness of fit chi-square = 1276.48 ,df=93) , p-value < 0.01) ค่า AIC เท่ากับ 17.95 ผลการวิเคราะห์ส่วนที่เหลือมีค่าเพียร์สัน (Pearson Chi-square= 1331.80,df=93) , p-value < 0.01 การตรวจสอบ Overdispersion จากสัดส่วนระหว่างค่าส่วนที่เหลือของเพียร์สัน (Pearson Residual) กับองศาความเป็นอิสระของความคลาดเคลื่อน (Residual degree of freedom) มีค่าเท่ากับ 14.32 และจากการวิเคราะห์ด้วยสถิติ

สกอร์ (score test) พบว่า Q_1 = 72.58 สถิติไคสแควร์ เท่ากับ 5267.96 (P-value < 0.001) Q_2 = 216.11 สถิติไคสแควร์ เท่ากับ 5267.96 (P-value < 0.001) แสดงว่า เกิดการกระจายที่มากเกินไป หรือ ตัวแปรตามมีความแปรปรวนมากกว่าค่าเฉลี่ย (Overdispersion) ซึ่งแสดงให้เห็นว่า ตัวแบบไม่มีความเหมาะสม ดังนั้นในการแก้ไขปัญหาการเกิดการกระจายที่มากเกินไป จึงต้องวิเคราะห์ด้วยตัวแบบทวินามเชิงลบ (Negative Binomial Model) เพื่อปรับให้ส่วนเบี่ยงเบนมาตรฐาน (Standard error: SE) น้อยลง (Hardin & Hilbe, 2001) โดยมีรูปแบบสมการทวินามเชิงลบ ดังนี้

$$\log(\mu_t) = -6.10 - (0.02 * t) - (0.09 * t_{max}) + \log(pop)$$

โดยที่

$\log(\mu_t)$ = อัตราการเกิดโรคมาลาเรีย

t = เวลา เป็นตัวแปรต่อเนื่อง

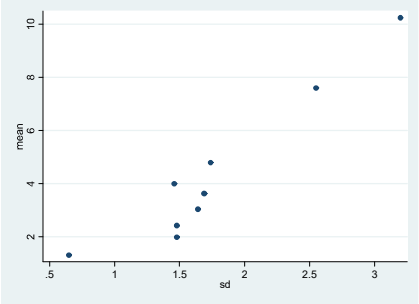
t_{max} = อุณหภูมิสูงสุด เป็นตัวแปรต่อเนื่อง (: องศาเซลเซียส)

pop = จำนวนประชากรกลางปี

จากการประเมินตัวแบบการพยากรณ์ ด้วยสถิติ Wald test พบว่า ตัวแปรเวลา มีความสัมพันธ์กับการเกิดโรคมาลาเรียที่ระดับนัยสำคัญ 0.05 (p-value < 0.05) ตัวแปรอุณหภูมิสูงสุดไม่มีความสัมพันธ์กับการเกิดโรคมาลาเรียที่ระดับนัยสำคัญ 0.05 (p-value > 0.05) แต่เมื่อทดสอบความสัมพันธ์ระหว่างตัวแปรทำนายโดยรวมกับการเกิดโรคมาลาเรีย โดยสถิติ Likelihood Ratio test พบว่า โดยรวมแล้วตัวแบบสามารถอธิบายถึงความสัมพันธ์ระหว่างตัวแปรทำนายกับการเกิดโรคมาลาเรียได้ที่ระดับนัยสำคัญ 0.05 (LR chi-square,df=2 เท่ากับ 75.88, p-value < 0.01) การตรวจสอบความเหมาะสม (Goodness of fit) ของตัวแบบพยากรณ์การเกิดโรคมาลาเรียจากค่า Deviance พบว่า ตัวแบบการทำนายมีความเหมาะสมของรูปแบบที่ระดับนัยสำคัญ 0.05 (Goodness of fit chi-square = 27.28, df=104, p-value > 0.05) จากค่า AIC เท่ากับ 10.35 ผลการวิเคราะห์ส่วนที่เหลือของเพียร์สัน Pearson Chi-square เท่ากับ 28.91,df=104 , p-value > 0.05

ตัวแบบ ARIMA Model

การทดสอบว่าอนุกรมเวลาของอัตราป่วยต่อแสนประชากรของการเกิดโรคมาลาเรีย $\{Y_t\}$ มีค่าเฉลี่ยและความแปรปรวนคงที่ พิจารณาจากแผนภาพการกระจายของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานดังรูปภาพที่ 2

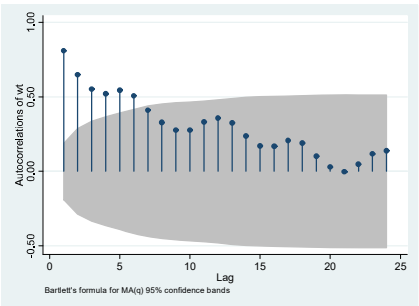


รูปภาพที่ 2 การกระจายของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ $\{Y_t\}$

จากรูปภาพที่ 2 พบว่า ค่าเฉลี่ยและความแปรปรวนไม่คงที่ จึงทำการแปลงข้อมูล ดังนี้

$$\{W_t\} = \log\{Y_t\}$$

และเมื่อพิจารณาคอเรลโรแกรม (correlogram) ของ $\{W_t\}$ ดังรูปภาพที่ 3

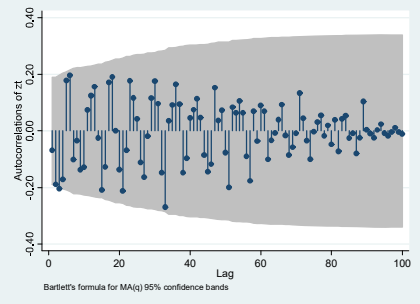


รูปภาพที่ 3 คอเรลโรแกรม ค่า r_k ของ $\{W_t\}$

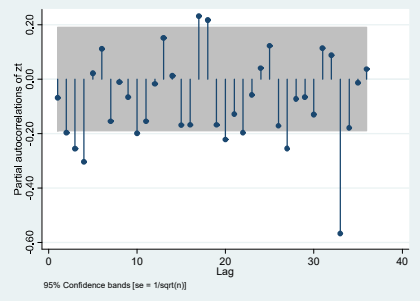
พบว่า อนุกรมเวลา $\{W_t\}$ มีแนวโน้มเป็นองค์ประกอบแสดงว่า อนุกรมเวลาชุดนี้มีค่าเฉลี่ยไม่คงที่ จึงทำการหาผลต่างของอนุกรมเวลา $\{W_t\}$ ดังนี้

$$\{Z_t\} = W_t - W_{t-1}$$

แล้วพิจารณาคอเรลโรแกรมของค่า r_k และ r_{kk} ของ {Zt} จากรูปภาพที่ 4 และรูปภาพที่ 5



รูปภาพที่ 4 คอเรลโรแกรมค่า r_k ของ $\{Z_t\}$



รูปภาพที่ 5 คอเรลโรแกรมค่า r_{kk} ของ $\{Z_t\}$

เมื่อพิจารณาคอเรลโรแกรมค่า r_{kk} ของอนุกรมเวลา Z_t ค่า r_{11} มีค่าที่ต่ำมากและไม่สามารถระบุตัวแบบได้ ในกรณีเช่นนี้ จึงกลับไปพิจารณา ACF และ PACF ของข้อมูลชุดเดิม คือ $\{W_t\}$ และโดยข้อเสนอแนะของ Wei(1989) จึงสามารถกำหนดตัวแบบให้แก่อนุกรมเวลาชุดนี้ ได้เป็น ARIMA(1,0,0) ดังนี้

$$\hat{W}_t = \quad 0.21 + 0.83(W_{t-1})$$

จากการตรวจสอบความเหมาะสมของตัวแบบการพยากรณ์พบว่า อนุกรมเวลาชุดนี้มีค่าเฉลี่ยและความแปรปรวนไม่คงที่สำหรับแต่ละค่าของ t ค่าสถิติทดสอบ Box-Ljung (Q_m) คอเรลโรแกรมของ r_k $\{e_t\}$ พบว่า r_k $|\{e_t\}|$ มีค่าไม่ต่างจาก 0 ค่าเฉลี่ยของความคลาดเคลื่อนเท่ากับ 0.32 มีค่าสถิติ t เท่ากับ 1.60 แสดงว่า ค่าเฉลี่ยของความคลาดเคลื่อนมีค่าแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (p-value >0.05) การแจกแจงของอนุกรมเวลาความคลาดเคลื่อนมีการแจกแจงที่เป็นปกติหรือค่าความแปรปรวนคงที่ (Z =0.99 , p-value>0.16)

การเปรียบเทียบตัวแบบการพยากรณ์

พิจารณาจากค่าวัดความถูกต้อง ได้แก่ ค่า MAD MSE และ MAPE เป็นเกณฑ์ในการเปรียบเทียบจำนวนผู้ป่วยโรคมาลาเรีย โดยตัวแบบ ARIMA(1,0,0) ต้องทำการแปลงค่าอัตราป่วยต่อแสนประชากร ให้เป็นจำนวนผู้ป่วยโรคมาลาเรียดังสมการ

$$pt = (\hat{Y}_t * \text{population})/100{,}000$$

โดยที่

pt คือ จำนวนผู้ป่วยโรคมาลาเรีย

$\hat{Y}_t$ คือ ค่าพยากรณ์อัตราป่วยต่อแสนประชากร ณ เวลา t

Population คือ จำนวนประชากรกลางปี

จากข้อมูลการเกิดโรคมาลาเรียระหว่างเดือนมกราคม 2543 ถึงธันวาคม 2551 จำนวน 108 ตัวอย่าง พบว่าการพยากรณ์ด้วยตัวแบบเชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยทวินามเชิงลบ (GLMs: Negative binomial regression) จะให้ค่า MAD MSE และ MAPE ต่ำที่สุด คือ มีค่า MAD เท่ากับ 12.33 ค่า MSE เท่ากับ 288.53 และ MAPE เท่ากับ 24.56 ซึ่งวัดความถูกต้องของการพยากรณ์ได้ 75.46%

เมื่อเปรียบเทียบค่าพยากรณ์ล่วงหน้า 12 เดือนกับค่าการเกิดโรคมาลาเรีย (ค่าสังเกต) ตั้งแต่เดือนมกราคม พ.ศ. 2552 ถึง เดือนธันวาคม พ.ศ. 2552 พบว่า การพยากรณ์ด้วยตัวแบบเชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยทวินามเชิงลบ (GLMs: Negative binomial regression) มีความเหมาะสมที่สุดในการพยากรณ์ คือ มีค่า MAD เท่ากับ 19.04 ค่า MSE เท่ากับ 558.22 และ MAPE เท่ากับ 33.46 ซึ่งวัดความถูกต้องของการพยากรณ์ได้ 66.54% (ดังตารางที่ 1 และตารางที่ 2)

บทสรุปและอภิปรายผล

จากผลการศึกษา การหาตัวแบบที่เหมาะสมในการพยากรณ์การเกิดโรคมาลาเรียในจังหวัดอุบลราชธานี เมื่อเปรียบเทียบจากค่า MAD MSE และ MAPE พบว่าตัวแบบเชิงเส้นโดยนัยทั่วไปประเภทสมการถดถอยทวินามเชิงลบ (GLMs: Negative binomial regression) มีความเหมาะสมที่สุดในการพยากรณ์ โดยมีค่า MAD MSE และ MAPE ต่ำที่สุดทั้งในขั้นตอนการพัฒนาตัวแบบพยากรณ์และการพยากรณ์ค่าล่วงหน้า ซึ่งสอดคล้องกับผลการศึกษาของ

Sriwattanapongse W., Kuning M. & Jansakul N. (2008) ในการศึกษาการเกิดโรคมาลาเรียในภาคตะวันตกเฉียงเหนือของประเทศไทย คือพบว่า เมื่อวิเคราะห์ด้วยสมการถดถอยปัวส์ซอง (GLMs: Poisson regression) แล้วเกิดปัญหา Overdispersion วิธีแก้ไขคือ การใช้สมการถดถอยทวินามเชิงลบ (GLMs: Negative binomial regression) ทำให้ได้ตัวแบบที่เหมาะสมในการพยากรณ์และทำให้ผลการวิเคราะห์ข้อมูลมีความถูกต้อง คือ จากการศึกษานี้เมื่อวิเคราะห์ด้วยตัวแบบ GLMs: Poisson Regression พบว่า ตัวแปรฤดูกาลมีความสัมพันธ์กับการเกิดโรคมาลาเรีย ซึ่งสอดคล้องกับทางทฤษฎี แต่ไม่สอดคล้องกับผลการศึกษาด้วยวิธีของบ็อกซ์และเจนกิ้นส์ ซึ่งพบว่า ฤดูกาลไม่มีความสัมพันธ์กับการเกิดโรคมาลาเรียในจังหวัดอุบลราชธานี และเมื่อประเมินตัวแบบ GLMs: Poisson Regression แล้วพบว่า ตัวแบบไม่มีความเหมาะสมเนื่องจากเมื่อพิจารณาค่า Deviance พบว่า ตัวแบบการทำนายไม่มีความเหมาะสมของรูปแบบที่ระดับนัยสำคัญ 0.05 (Goodness of fit chi-square = 1276.48, df=93, p-value < 0.01) ค่า AIC เท่ากับ 17.95 ผลการวิเคราะห์ส่วนที่เหลือของเพียร์สัน Pearson Chi-square เท่ากับ 1331.80, df=93, p-value < 0.01 การตรวจสอบ Overdispersion จากสัดส่วนระหว่างค่าส่วนที่เหลือของเพียร์สัน (Pearson Residual) กับองศาความเป็นอิสระของค่าส่วนที่เหลือ (Residual degree of freedom) มีค่าเท่ากับ 14.32 และจากการวิเคราะห์ด้วยสถิติสกอร์ (score test) พบว่า Q_1 เท่ากับ 72.58 สถิติไคสแควร์ เท่ากับ 5267.96 (p-value < 0.001) Q_2 เท่ากับ 216.11 สถิติไคสแควร์ เท่ากับ 5267.96 (p-value < 0.001) แสดงว่า เกิดปัญหาการกระจายที่มากเกินไป (Overdispersion) แต่เมื่อแก้ไขปัญหา Overdispersion โดยการวิเคราะห์ด้วยตัวแบบถดถอยทวินามเชิงลบ พบว่า ฤดูกาลไม่มีความสัมพันธ์กับการเกิดโรคมาลาเรีย ซึ่งสอดคล้องกับการวิเคราะห์ด้วยวิธีบ็อกซ์และเจนกิ้นส์ ซึ่งส่งให้ผลการวิเคราะห์สามารถพยากรณ์ค่าได้อย่างแม่นยำและถูกต้องมีค่าความคลาดเคลื่อนต่ำที่สุด

ข้อเสนอแนะจากการวิจัย

ควรมีการนำเทคนิคการพยากรณ์ไปเผยแพร่และนำไปประยุกต์ใช้จริงในการกำหนดเป้าหมายในการป้องกันโรคระบาด เพื่อความถูกต้องและแม่นยำในการวางแผนเพื่อเป็นประโยชน์ต่องานด้านสาธารณสุข การศึกษาด้วยวิธีของบ็อกซ์และเจนกิ้นส์ในอนาคต ควรศึกษาวิธีการพยากรณ์ที่ใช้ปัจจัยร่วมในการพยากรณ์นอกเหนือจากข้อมูลในอดีตเพียงอย่างเดียว และนำไปขยายผลในการพยากรณ์โรคที่เป็นปัญหาสาธารณสุขปัจจุบันโรคอื่น ๆ เช่น โรคไข้เลือดออก และควรขยายพื้นที่การศึกษาให้ครอบคลุมพื้นที่สาธารณสุขเขต 7

กิตติกรรมประกาศ

ขอขอบคุณรองศาสตราจารย์ ดร.วัฒนาวดี ศรีวัฒนพงศ์ คณาจารย์ภาควิชาชีวสถิติและประชากรศาสตร์ มหาวิทยาลัยขอนแก่นทุกท่าน ที่กรุณาให้ข้อเสนอที่เป็นประโยชน์ต่อการศึกษาครั้งนี้

เอกสารอ้างอิง

บุญส่ง สอนส่งเสริม. (2553). **มาลาเรีย.** ค้นเมื่อ 16 เมษายน 2553 , จาก http://student.nu.ac.th/Malaria/1html.

สำนักโรคติดต่อนำโดยแมลง. (2552). **สรุปรายงานการเฝ้าระวังโรค.** ค้นเมื่อ 15 เมษายน 2553 , จาก http://www.thaivbd.org/cms/index.php.

สำนักงานควบคุมและป้องกันโรคที่ 7. (2552).**วิเคราะห์สถานการณ์โรครายเดือนย้อนหลัง.** ค้นเมื่อ 15 เมษายน 2553, จาก http://vbd.dpc7.net/content/blogcategory/14/34/.

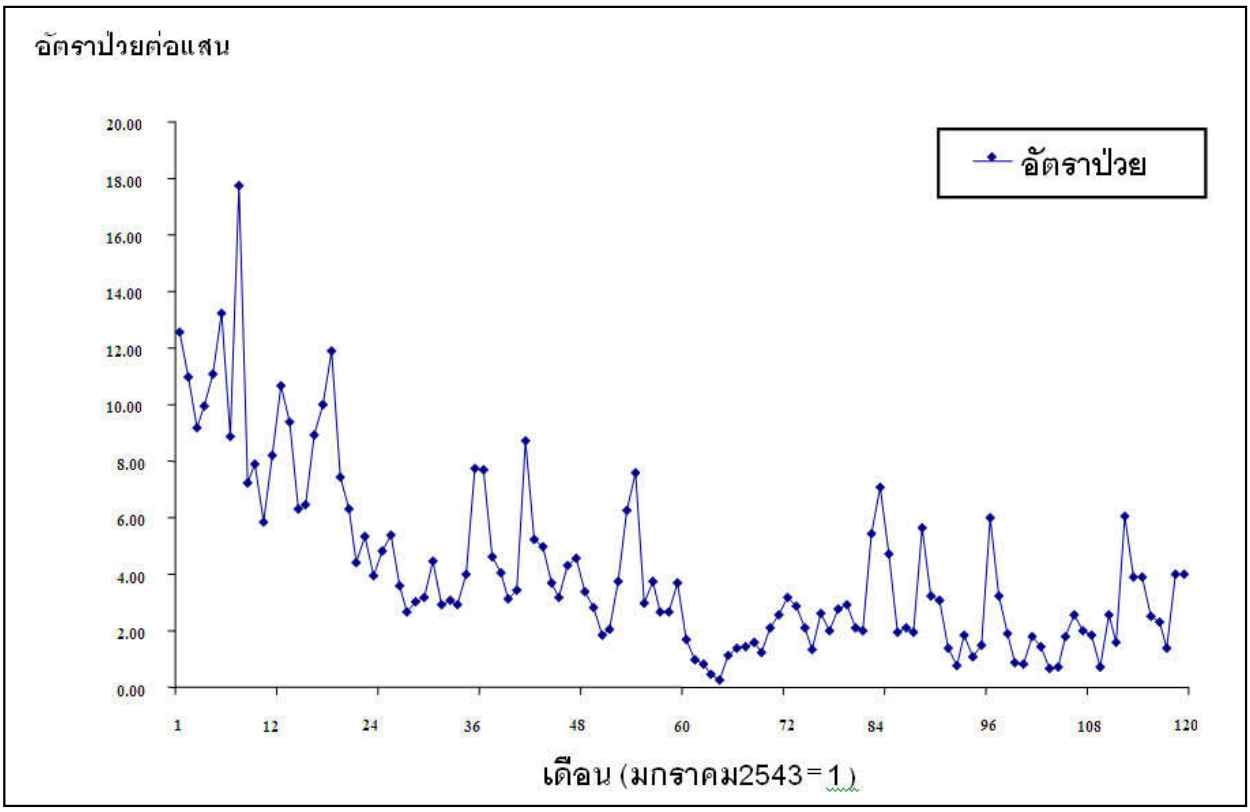
ศูนย์อุตุนิยมวิทยาภาคตะวันออกเฉียงเหนือ. (2552). **พยากรณ์อากาศประจำเดือน.** ค้นเมื่อ 16 เมษายน 2553 , จาก http://www.ubonmet.org/, 2545.

Gaudart, J., et al. (2009). Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali. **Malaria Journal,** 8(61), 1-16.

Hardin, J., & Hilbe, J. (2001). **Generalized Linear Models and Extensions**. Texas: Stata corporation.

Sriwattanapongse W., Kuning M. & Jansakul N. (2008). Malaria in North - Western Thailand. **Songklanakarin Journal of Science and Technology**, 30(2). 207-209.

Wei, WS. (1989).**Time Series Analysis: Univariate and Multivariate Methods.** New York: Addison-Wesley.



รูปภาพที่ 1 อัตราป่วยต่อแสนประชากรของโรคมาลาเรีย จังหวัดอุบลราชธานี ระหว่าง มกราคม 2543-ธันวาคม 2552

ตารางที่ 1 เปรียบเทียบค่า MAD, MSE และ MAPE ของอนุกรมเวลา การเกิดโรคมาลาเรียรายเดือน จำนวน 108 เดือน

ค่าวัดความถูกต้อง	ตัวแบบ GLM Poisson	ตัวแบบ GLM Negative Binomial	ตัวแบบ ARIMA(1,0,0)
MAD	21.29	12.33	23.28
MSE	1018.61	288.53	1179.30
MAPE	51.33	24.56	38.33

ตารางที่ 2 เปรียบเทียบค่า MAD, MSE และ MAPE ของอนุกรมเวลาการเกิดโรคมาลาเรียรายเดือน ล่วงหน้า 12 เดือน

ค่าวัดความถูกต้อง	ตัวแบบ GLM Poisson	ตัวแบบ GLM Negative Binomial	ตัวแบบ ARIMA(1,0,0)
MAD	30.58	19.04	19.45
MSE	1464.83	558.22	590.09
MAPE	53.75	33.46	53.00