

การเปรียบเทียบความแม่นยำของการพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี

A COMPARATIVE PREDICTION ACCURACY OF BIRTH NUMBER OF THE POPULATION PATHUM THANI PROVINCE

คชินทร์ โกขุนทดารณ์*

*Kachin Goganutapon**

สาขาวิชาคณิตศาสตร์ประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏวไลยอลงกรณ์
ในพระบรมราชูปถัมภ์

*Program in Applied Mathematics, Faculty of Science and Technology, Valaya Alongkorn Rajabhat
University under the Royal Patronage.*

**Corresponding author, e-mail: kachin@vru.ac.th*

Received: 11 February 2020; **Revised:** 6 November 2020; **Accepted:** 31 March 2021

บทคัดย่อ

การวิจัยนี้ศึกษาเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์สำหรับพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี โดยใช้ข้อมูลจากเว็บไซต์ของระบบสถิติทางการทะเบียน สำนักบริหารการทะเบียน กรมการปกครอง ตั้งแต่ ปี พ.ศ. 2536 ถึงปี พ.ศ. 2562 จำนวน 27 ค่า ข้อมูลแบ่งออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่ปี พ.ศ. 2536 ถึงปี พ.ศ. 2559 จำนวน 24 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วย วิธีบอกซ์-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลท์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉก และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ ข้อมูลชุดที่ 2 ตั้งแต่ปี พ.ศ. 2560 ถึง ปี พ.ศ. 2562 จำนวน 3 ค่า สำหรับการเปรียบเทียบความแม่นยำของวิธีการพยากรณ์โดยใช้เกณฑ์ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ที่ต่ำที่สุด ผลการวิจัยพบว่า วิธีการพยากรณ์ที่มีความแม่นยำที่สุดคือ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์

คำสำคัญ: จำนวนการเกิดของประชากร วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

Abstract

This research is aimed to compare the prediction accuracy of forecasting model for the birth number of the population Pathum Thani Province. The data gathered from the website of Registration

system - Office of Registration Administration, Department of Local Administration during year 1993 to the year 2019 of 27 values which were used and separated into 2 groups. The first group contained 24 values from year 1993 to the year 2016 for comparing and searching for forecasting models. There were forecasting methods: Box-Jenkins, Holt's exponential smoothing method, Damped trend exponential smoothing method and Brown's exponential smoothing method. The second set had 3 values from year 2017 to the year 2019 for comparing and finding the most suitable forecasting method via criteria of the lowest root mean squared error (RMSE) and mean absolute percentage error (MAPE). The result indicated Brown's exponential smoothing method was the best method.

Keywords: Birth number of the population, Brown's exponential smoothing method, Mean absolute percentage error

บทนำ

นับตั้งแต่ปี พ.ศ. 2561 แนวโน้มจำนวนของประชากรเริ่มลดลง ซึ่งเห็นได้จากบทความของ UNFPA ที่ได้กล่าวว่า ในภูมิภาคเอเชียและแปซิฟิกมีขนาดครอบครัวที่เล็กลง สะท้อนให้เห็นถึงการตัดสินใจของผู้คนเกี่ยวกับอนามัยเจริญพันธุ์ของตนเองว่าจะมีลูกน้อยหรือมากตามที่แต่ละคนต้องการและตามช่วงเวลาที่ต้องการ ขนาดของครอบครัวยังมีความสัมพันธ์กับสิทธิอนามัยการเจริญพันธุ์ ซึ่งมีความเกี่ยวเนื่องกับสิทธิด้านอื่น ๆ ที่แต่ละปัจเจกควรจะได้รับ ได้แก่ สิทธิที่จะได้รับการดูแลด้านสุขภาพ การศึกษา และการจ้างงาน เมื่อใดที่ผู้คนได้รับสิทธิก็มีแนวโน้มว่าพวกเขาจะสร้างความพึงพอใจ แต่เมื่อใดที่ถูกจำกัดผู้คนก็จะไม่สามารถพัฒนาศักยภาพของตนเองได้ ซึ่งเป็นอุปสรรคต่อการพัฒนาทั้งทางเศรษฐกิจและสังคม [1] นอกจากนี้ยังพบว่า อัตราการคลอดบุตรในวัยรุ่นไทย พ.ศ. 2561 ลดลงจาก พ.ศ. 2560 ทั้งในกลุ่มอายุ 10-14 ปี และ 15-19 ปี เท่ากับ 1.2 ต่อประชากรหญิงอายุ 10-14 ปี 1,000 คน และ 35.0 ต่อประชากรหญิงอายุ 15-19 ปี 1,000 คน ตามลำดับ [2] ผู้วิจัยได้สืบค้นงานวิจัยที่เกี่ยวกับการพยากรณ์จำนวนการเกิดของประชากรโดยวิธีทางสถิติ ซึ่ง Antonio Caleiro ได้ศึกษาการคาดการณ์จำนวนการเกิดในโปรตุเกส ใช้ข้อมูลจำนวนการเกิดเป็นรายเดือน ตั้งแต่เดือน มกราคม พ.ศ. 2520 ถึง ธันวาคม พ.ศ. 2551 โดยใช้แบบจำลอง ARIMA [3] ผู้วิจัยจึงมีความสนใจที่จะศึกษาการสร้างตัวแบบพยากรณ์ทางสถิติ จำนวน 4 วิธี คือ วิธีบอซ-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลท์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแดม และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ จากนั้นจึงเลือกวิธีที่มีความแม่นยำมากที่สุด 1 วิธี โดยการพิจารณาจากเกณฑ์ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ที่ต่ำที่สุด เพื่อให้ได้ตัวแบบพยากรณ์ที่มีความแม่นยำมากที่สุด สำหรับพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี เพื่อเป็นแนวทางในการเตรียมความพร้อมในการพัฒนา จังหวัดปทุมธานี ในอนาคตต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ จำนวนการเกิดของประชากร จังหวัดปทุมธานี

วิธีดำเนินการวิจัย

การสร้างตัวแบบพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี ใช้ข้อมูลจากเว็บไซต์ระบบสถิติทางการทะเบียน สำนักบริหารการทะเบียน กรมการปกครอง [4] โดยเริ่มใช้ข้อมูลรายปีตั้งแต่ปี พ.ศ. 2536 ถึง ปี พ.ศ. 2562 จำนวน 27 ค่า โดยแบ่งข้อมูลออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่เดือน ปี พ.ศ. 2553 ถึง ปี พ.ศ. 2559 จำนวน 24 ค่า สำหรับเปรียบเทียบหาตัวแบบพยากรณ์ โดยงานวิจัยนี้ใช้วิธีการพยากรณ์คือ วิธีบ็อกซ์-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังแบบโฮลท์วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉกและวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ และข้อมูลชุดที่ 2 ตั้งแต่ปี พ.ศ. 2560 ถึง ปี พ.ศ. 2562 จำนวน 3 ค่า สำหรับการเปรียบเทียบหาวิธีการพยากรณ์ที่มีความแม่นยำโดยใช้เกณฑ์ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ที่ต่ำที่สุด

ศึกษาลักษณะการเคลื่อนไหวของอนุกรมเวลา

การศึกษาลักษณะการเคลื่อนไหวของอนุกรมเวลาเป็นการพิจารณาในเบื้องต้นว่าอนุกรมเวลามีการเปลี่ยนแปลงในลักษณะใด มีส่วนประกอบของอนุกรมเวลาใดบ้าง (แนวโน้ม ฤดูกาล วัฏจักร หรือเหตุการณ์ที่ผิดปกติ) โดยพิจารณาจากแผนภาพการกระจายของอนุกรมเวลาเทียบกับเวลา เพื่อความเหมาะสมของการเลือกใช้วิธีการทางสถิติสำหรับการวิเคราะห์ข้อมูลต่อไป [5]

การวิเคราะห์ข้อมูลทางสถิติ

วิธีการสร้างตัวแบบพยากรณ์ในการวิจัยครั้งนี้ได้ศึกษา 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังแบบโฮลท์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉก และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ ดังแสดงรายละเอียดต่อไปนี้

การพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์ (Box-Jenkins Method)

สำหรับการพยากรณ์ระยะสั้น วิธีบ็อกซ์-เจนกินส์เป็นวิธีการพยากรณ์ที่มีความถูกต้องสูงกว่าวิธีอื่น ๆ [6] เนื่องจากการกำหนดตัวแบบโดยตรวจสอบคุณสมบัติฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) ซึ่งพิจารณาภายใต้อนุกรมเวลาที่เป็นสเตชันนารี (Stationary Time Series) หรืออนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่ โดยกรณีที่อนุกรมเวลาไม่เป็นสเตชันนารี ต้องแปลงอนุกรมเวลาให้เป็นสเตชันนารีก่อนที่จะกำหนดตัวแบบ เช่น การแปลงข้อมูลด้วยการหาผลต่างหรือผลต่างฤดูกาล (Difference or Seasonal Difference) กรณีค่าเฉลี่ยไม่คงที่การแปลงข้อมูลด้วยลอการิทึมสามัญหรือลอการิทึมธรรมชาติ (Logarithm or Natural Logarithm) การแปลงข้อมูลด้วยเลขยกกำลัง เช่น ยกกำลัง 0.5 (Square Root Transformation) หรือยกกำลัง 2 (Square Transformation) กรณีความแปรปรวนไม่คงที่ เป็นต้น ตัวแบบทั่วไปของวิธีบ็อกซ์-เจนกินส์ คือ Seasonal Autoregressive Integrated Moving Average: SARIMA(p,d,q)(P,D,Q)_s แสดงดังสมการ (1) และขั้นตอนการสร้างตัวแบบพยากรณ์ [5]

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

เมื่อ Y_t แทนอนุกรมเวลา ณ เวลา t

ε_t แทนอนุกรมเวลาของความคลาดเคลื่อนที่มีการแจกแจงปกติและเป็นอิสระกัน ด้วยค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่ทุกช่วงเวลา

$\delta = \mu\phi_p(B)\Phi_P(B^s)$ แทนค่าคงที่ โดยที่ μ แทนค่าเฉลี่ยของอนุกรมเวลาที่เป็นสเตชันนารี

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ แทนตัวดำเนินการสหสัมพันธ์ในตัวเองแบบไม่มีฤดูกาล
อันดับที่ p (Non-Seasonal Autoregressive Operator of Order p : AR(p))

$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$ แทนตัวดำเนินการสหสัมพันธ์ในตัวเองแบบมีฤดูกาล
อันดับที่ P (Seasonal Autoregressive Operator of Order p : SAR(P))

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ แทนตัวแบบดำเนินการเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับที่
 q (Non-Seasonal Moving Average Operator of Order q : MA(q))

$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}$ แทนตัวแบบดำเนินการเฉลี่ยเคลื่อนที่แบบมีฤดูกาล
อันดับที่ Q (Seasonal Moving Average Operator of Order Q : SMA(Q))

เมื่อ t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n โดยที่ n แทนจำนวนข้อมูลในอนุกรมเวลาชุดที่ 1

s แทนจำนวนฤดูกาล

d และ D แทนลำดับที่ของการหาผลต่างและผลต่างฤดูกาล ตามลำดับ

B แทนตัวดำเนินการถอยหลัง (Backward Operator) โดยที่ $B^s Y_t = Y_{ts}$

การพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์ (Holt's exponential smoothing method: Holt)

การทำให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์ มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้มเป็นเส้นตรงและไม่มีส่วนประกอบของความผันแปรตามฤดูกาล มีค่าคงที่การทำให้เรียบ 2 ตัว คือ ค่าคงที่การปรับเรียบของค่าระดับ (Level: α) และค่าคงที่การปรับเรียบของค่าความชัน (Trend: γ) ตัวแบบเขียนได้ดังสมการที่ (2) และตัวแบบพยากรณ์เขียนได้ดังสมการที่ (3) [6]

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (2)$$

$$\hat{Y}_{t+m} = a_t + b_t(m) \quad (3)$$

เมื่อ Y_t แทนอนุกรมเวลา ณ เวลา t , β_0 และ β_1 แทนพารามิเตอร์ของตัวแบบแสดงระยะตัดแกนและความชันของแนวโน้ม ตามลำดับ ε_t แทน อนุกรมเวลาของความคลาดเคลื่อนที่มีการแจกแจงปกติและเป็นอิสระกัน ด้วยค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ทุกช่วงเวลา $\hat{Y}_{t+m}$ แทนค่าพยากรณ์ ณ เวลา $t + m$ โดยที่ m แทนจำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า a_t และ b_t แทนค่าประมาณ ณ เวลา t ของพารามิเตอร์ β_0 และ β_1 ตามลำดับ โดยที่ $a_t = \alpha Y_t + (1 - \alpha)(a_{t-1} + b_{t-1})$, $b_t = \gamma(a_t - a_{t-1}) + (1 - \gamma)b_{t-1}$, α และ γ แทน ค่าคงที่การทำให้เรียบ โดยที่ $0 < \alpha < 1$ และ $0 < \gamma < 1$, t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n โดยที่ n แทนจำนวนข้อมูลในอนุกรมเวลาชุดที่ 1

การสร้างตัวแบบพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแดม (Damped trend exponential smoothing method: Damped)

วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแดม มีความเหมาะสมกับอนุกรมเวลาที่ไม่มีส่วนประกอบของความผันแปรตามฤดูกาล และมีอัตราการเพิ่มขึ้น (หรือลดลง) เปลี่ยนแปลงช้ากว่าการเพิ่มขึ้น (หรือลดลง) ของแนวโน้มที่เป็นเส้นตรง มีค่าคงที่การทำให้เรียบ 3 ตัว คือ ค่าคงที่การปรับเรียบของค่าระดับ (α) ค่าคงที่การปรับเรียบของค่าความชัน (γ) และค่าคงที่การปรับเรียบของค่าความชันแบบแดม (Damped Trend) (Φ) ตัวแบบพยากรณ์เขียนได้ดังนี้ [7]

$$\hat{Y}_{t+m} = a_t + b_t \sum_{i=1}^m \phi^i \quad (4)$$

เมื่อ $\hat{Y}_{t+m}$ แทนค่าพยากรณ์ ณ เวลา $t + m$ โดยที่ m แทนจำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า a_t และ b_t แทนค่าประมาณ ณ เวลา t ของพารามิเตอร์ β_0 และ β_1 ตามลำดับ

โดยที่ $a_t = \alpha Y_t + (1 - \alpha)(a_{t-1} + \phi b_{t-1})$, $b_t = \gamma(a_t - a_{t-1}) + (1 - \gamma)\phi b_{t-1}$, α, γ และ ϕ แทนค่าคงที่การทำให้เรียบโดยที่ $0 < \alpha < 1$, $0 < \gamma < 1$ และ $0 < \phi < 1$, t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n , n แทนจำนวนข้อมูลในอนุกรมเวลาชุดที่ 1

การสร้างตัวแบบพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ (Brown's exponential smoothing method: Brown)

การทำให้เรียบแบบเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้มเป็นเส้นตรงและไม่มีส่วนประกอบของความผันแปรตามฤดูกาล เช่นเดียวกับการทำให้เรียบแบบเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์ แต่มีการกำหนดให้ค่าคงที่การปรับเรียบของค่าระดับและค่าคงที่การปรับเรียบของค่าความชันเท่ากัน ดังนั้นอาจกล่าวได้ว่าการทำให้เรียบแบบเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์เป็นกรณีพิเศษของการทำให้เรียบแบบเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์ ตัวแบบพยากรณ์เขียนได้ดังนี้ [7]

$$\hat{Y}_{t+m} = a_t + b_t \left[(m - 1) + \frac{1}{\alpha} \right] \quad (5)$$

เมื่อ $\hat{Y}_{t+m}$ แทนค่าพยากรณ์ ณ เวลา $t + m$ โดยที่ m แทนจำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า a_t และ b_t แทน ค่าประมาณ ณ เวลา t ของพารามิเตอร์ β_0 และ β_1 ตามลำดับ โดยที่ $a_t = \alpha Y_t + (1 - \alpha)a_{t-1}$, $b_t = \alpha(a_t - a_{t-1}) + (1 - \alpha)b_{t-1}$, α แทนค่าคงที่การทำให้เรียบ โดยที่ $0 < \alpha < 1$, t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n , n แทนจำนวนข้อมูลในอนุกรมเวลาชุดที่ 1

การเปรียบเทียบความแม่นยำของค่าพยากรณ์

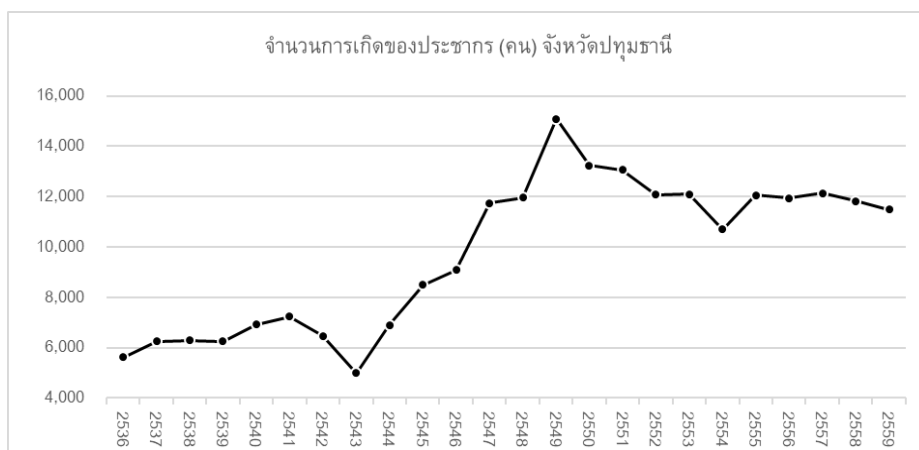
การวิจัยครั้งนี้ได้ศึกษาเปรียบเทียบตัวแบบพยากรณ์ที่มีความแม่นยำกับอนุกรมเวลา โดยการเปรียบเทียบค่ารากที่สองของคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ที่ต่ำที่สุดจากวิธีการพยากรณ์ทั้ง 4 วิธี ตัวแบบพยากรณ์ที่มีค่า RMSE และ MAPE ต่ำที่สุดจัดเป็นตัวแบบที่มีความแม่นยำมากที่สุด เนื่องจากให้ค่าพยากรณ์ที่มีความแตกต่างกับข้อมูลจริงน้อยที่สุด เกณฑ์ RMSE และ MAPE [6] ดังนี้

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \text{ และ } MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (6)$$

เมื่อ Y_t แทนค่าของข้อมูลจริง ณ เวลา t , $\hat{Y}_t$ แทน ค่าพยากรณ์ ณ เวลา t , n แทนจำนวนข้อมูล

ผลการวิจัย

การพิจารณาลักษณะการเคลื่อนไหวของอนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี ของข้อมูลชุดที่ 1 ตั้งแต่ปี พ.ศ. 2536 ถึง ปี พ.ศ. 2559 จำนวน 24 ค่า ดังภาพที่ 1 พบว่าจำนวนการเกิดของประชากร จังหวัดปทุมธานี มีส่วนประกอบของแนวโน้มช่วงปี พ.ศ. 2557 ถึง ปี พ.ศ. 2559 จำนวนการเกิดของประชากร จังหวัดปทุมธานี มีแนวโน้มลดลง

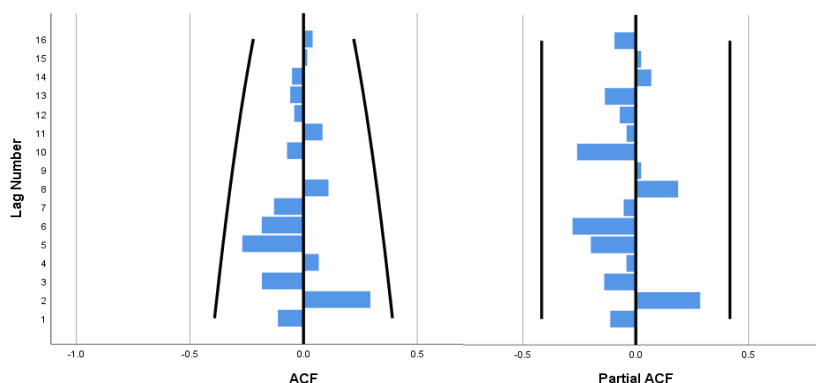


ภาพที่ 1 ลักษณะการเคลื่อนไหวของอนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี ตั้งแต่ปี พ.ศ. 2536 ถึง ปี พ.ศ. 2559

ที่มา: Goganutapon, K. (2020). *A comparative prediction accuracy of birth number of the population Pathum Thani Province*. n.p. [8]

ผลการพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์

เนื่องจากอนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี ชุดนี้มีส่วนประกอบของแนวโน้ม ผู้วิจัยจึงแปลงข้อมูลด้วยการหาผลต่างอันดับที่ 1 ($d = 1$) ได้กราฟ ACF และ PACF ได้ผลแสดงดังภาพที่ 2



ภาพที่ 2 กราฟ ACF และ PACF ของอนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี เมื่อหาผลต่างอันดับที่ 1

ที่มา: Goganutapon, K. (2020). *A comparative prediction accuracy of birth number of the population Pathum Thani Province*. n.p. [8]

จากภาพที่ 2 อนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี มีลักษณะคงที่ จึงสามารถกำหนดตัวแบบพยากรณ์ที่เป็นไปได้ พร้อมกับประมาณค่าพารามิเตอร์ โดยกำหนดรูปแบบเริ่มต้น คือ ตัวแบบ ARIMA(1,1,1) แสดงดังตารางที่ 1

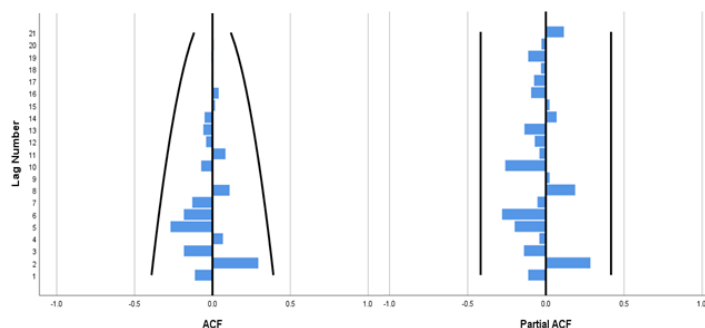
ตารางที่ 1 ค่าประมาณพารามิเตอร์ ค่า BIC และค่าสถิติ Ljung-Box Q ของตัวแบบ ARIMA(p,d,q)

ค่าประมาณพารามิเตอร์		ตัวแบบ ARIMA(p,d,q)			
		ARIMA(1,1,1)	ARIMA(0,1,1)	ARIMA(1,1,0)	ARIMA(0,1,0)
ค่าคงที่	ค่าประมาณ	252.24465	255.90496	255.91657	254.95652
	p-value	0.31275	0.31195	0.29794	0.33836
AR(1)	ค่าประมาณ	-0.82022		-0.10979	
	ϕ_1 p-value	0.08127		0.61974	
MA(1)	ค่าประมาณ	-0.67893	0.07331		
	θ_1 p-value	0.25773	0.74141		
BIC		14.69146	14.57156	14.56691	14.39702
Ljung-Box Q (ณ lag 18)		8.64234	9.35167	9.44322	9.50696
p-value		0.92741	0.92858	0.92532	0.94682

จากตารางที่ 1 ตัวแบบพยากรณ์ ARIMA(0,1,0) ที่มีค่า BIC ต่ำที่สุดคือ BIC = 14.39702 และมีค่าสถิติ Ljung-Box Q ไม่มีนัยสำคัญที่ระดับ 0.01 (Ljung-Box Q ณ lag 18 = 9.50696, df = 18, p-value = 0.94682) ค่า RMSE = 1262.992 และค่า MAPE = 9.8693 เมื่อตรวจสอบคุณลักษณะของความคลาดเคลื่อนจากการพยากรณ์ ที่ระดับนัยสำคัญ 0.01 พบว่าความคลาดเคลื่อนมีการแจกแจงแบบปรกติ (Shapiro-Wilk Z = 0.15, p-value = 0.40) มีค่าเฉลี่ยเท่ากับศูนย์ ($t = -0.00017$, p-value = 0.999) ความคลาดเคลื่อนมีการเคลื่อนไหวเป็นอิสระกัน แสดงรายละเอียดดังภาพที่ 3 ซึ่งพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในบางส่วนของความคลาดเคลื่อนตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 99 และมีความแปรปรวนเท่ากันทุกช่วงเวลา (Levene Statistic = 0.385, p-value = 0.546) ดังนั้นตัวแบบพยากรณ์ที่ได้มีความเหมาะสมซึ่งตัวแบบพยากรณ์ คือ

$$\hat{Y}_t = 254.95652 + Y_{t-1} \quad (7)$$

เมื่อ $\hat{Y}_t$ แทนค่าพยากรณ์ ณ เวลา t และ Y_{t-1} แทน อนุกรมเวลา ณ เวลา $t - 1$



ภาพที่ 3 กราฟ ACF และ PACF ของความคลาดเคลื่อนจากการพยากรณ์ โดยวิธีบอซ-เจนกินส์

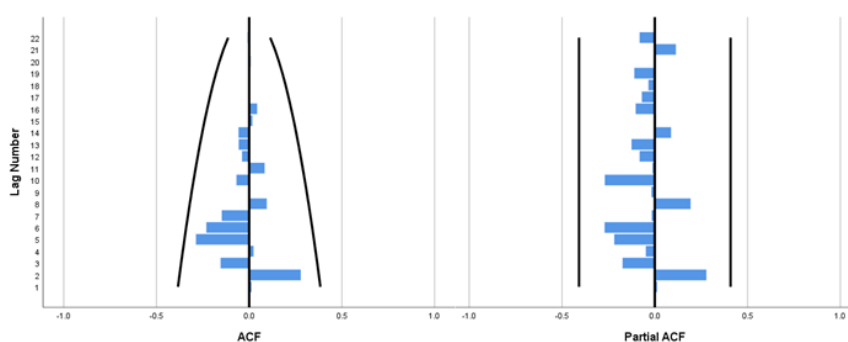
ที่มา: Goganutapon, K. (2020). A comparative prediction accuracy of birth number of the population Pathum Thani Province. n.p. [8]

ผลการพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์

จากการสร้างตัวแบบโดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์ พบว่า BIC มีค่าเท่ากับ 14.52454 และมีค่าสถิติ Ljung-Box Q ไม่มีนัยสำคัญที่ระดับ 0.01 (Ljung-Box Q ณ lag 18 = 9.79791, df = 16, p-value = 0.87697) ค่า RMSE = 1248.69279 และค่า MAPE = 9.68498 เมื่อตรวจสอบคุณลักษณะของความคลาดเคลื่อนจากการพยากรณ์ ที่ระดับนัยสำคัญ 0.01 พบว่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ (Shapiro-Wilk Z = 0.949, p-value = 0.263) มีค่าเฉลี่ยเท่ากับศูนย์ (t = -0.374, p-value = 0.712) ความคลาดเคลื่อนมีการเคลื่อนไหวเป็นอิสระกัน แสดงรายละเอียดดังภาพที่ 4 ซึ่งพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในบางส่วนของความคลาดเคลื่อนตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 99 และมีความแปรปรวนเท่ากันทุกช่วงเวลา (Levene Statistic = 0.03, p-value = 0.954) ดังนั้นตัวแบบพยากรณ์ที่ได้มีความเหมาะสม ตัวแบบพยากรณ์แสดงดังนี้

$$\hat{Y}_{t+m} = 11587.07349 + 630.789(m) \quad (8)$$

เมื่อ $\hat{Y}_{t+m}$ คือค่าพยากรณ์ ณ เวลา $t + m$ โดยที่ $m = 1$ แทนปี พ.ศ. 2560 $\alpha = 0.90092, \gamma = 0.00002$



ภาพที่ 4 กราฟ ACF และ PACF ของความคลาดเคลื่อนจากการพยากรณ์
วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์

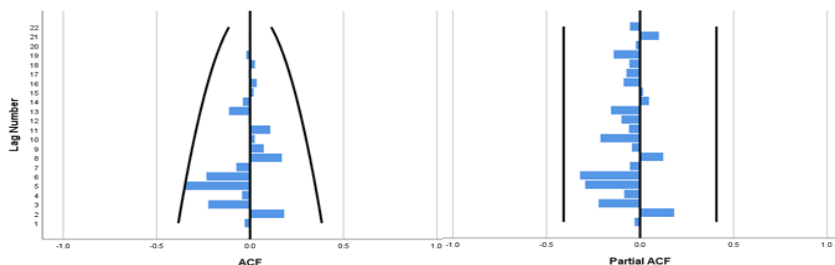
ที่มา: Goganutapon, K. (2020). *A comparative prediction accuracy of birth number of the population Pathum Thani Province*. n.p. [8]

ผลการพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์

จากการสร้างตัวแบบโดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ พบว่า BIC มีค่าเท่ากับ 14.50893 และมีค่าสถิติ Ljung-Box Q ไม่มีนัยสำคัญที่ระดับ 0.01 (Ljung-Box Q ณ lag 18 = 11.52831, df = 17, p-value = 0.82782) ค่า RMSE = 1323.79127 และค่า MAPE = 10.66905 เมื่อตรวจสอบคุณลักษณะของความคลาดเคลื่อนจากการพยากรณ์ ที่ระดับนัยสำคัญ 0.01 พบว่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ (Shapiro-Wilk Z = 0.952, p-value = 0.295) มีค่าเฉลี่ยเท่ากับศูนย์ (t = -0.200, p-value = 0.843) ความคลาดเคลื่อนมีการเคลื่อนไหวเป็นอิสระกัน แสดงรายละเอียดดังภาพที่ 5 ซึ่งพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในบางส่วนของความคลาดเคลื่อนตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 99 และมีความแปรปรวนเท่ากันทุกช่วงเวลา (Levene Statistic = 0.264, p-value = 0.616) ดังนั้นตัวแบบพยากรณ์ที่ได้มีความเหมาะสม ตัวแบบพยากรณ์แสดงดังนี้

$$\hat{Y}_{t+m} = 11664.24334 - 144.10576 \left[(m-1) + \frac{1}{0.56757} \right] \quad (9)$$

เมื่อ $\hat{Y}_{t+m}$ คือค่าพยากรณ์ ณ เวลา $t + m$ โดยที่ $m = 1$ แทนปี พ.ศ. 2560 ซึ่ง $\alpha = 0.56757$



ภาพที่ 5 กราฟ ACF และ PACF ของความคลาดเคลื่อนจากการพยากรณ์
วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์

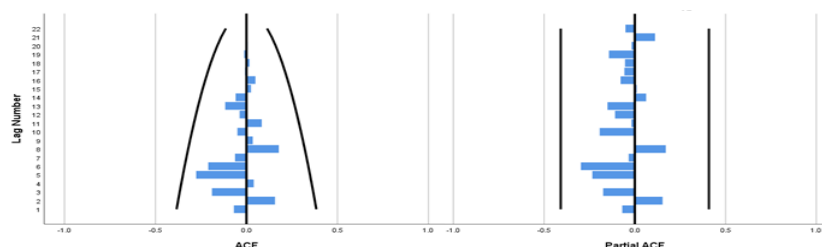
ที่มา: Goganutapon, K. (2020). *A comparative prediction accuracy of birth number of the population Pathum Thani Province*. n.p. [8]

ผลการพยากรณ์โดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแตรม

จากการสร้างตัวแบบโดยวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแตรม พบว่า BIC มีค่าเท่ากับ 14.68119 และมีค่าสถิติ Ljung-Box Q ไม่มีนัยสำคัญที่ระดับ 0.01 (Ljung-Box Q ณ lag 18 = 9.1906, df= 15, p-value = 0.86734) ค่า RMSE = 1263.91462 และค่า MAPE = 9.54358 เมื่อตรวจสอบคุณลักษณะของความคลาดเคลื่อนจากการพยากรณ์ ที่ระดับนัยสำคัญ 0.01 พบว่าความคลาดเคลื่อนมีการแจกแจงแบบปรกติ (Shapiro-Wilk Z = 0.959, p-value = 0.412) มีค่าเฉลี่ยเท่ากับศูนย์ (t = 0.674, p-value = 0.507) ความคลาดเคลื่อนมีการเคลื่อนไหวเป็นอิสระกัน แสดงรายละเอียดดังภาพที่ 6 ซึ่งพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในบางส่วนของความคลาดเคลื่อนตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 99 และมีความแปรปรวนเท่ากันทุกช่วงเวลา (Levene Statistic = 0.035, p-value = 0.854) ดังนั้นตัวแบบพยากรณ์ที่ได้มีความเหมาะสม ตัวแบบพยากรณ์แสดงดังนี้

$$\hat{Y}_{t+m} = 11653.70847 - 328.95096 \sum_{i=1}^m (0.52485)^i \quad (10)$$

เมื่อ $\hat{Y}_{t+m}$ คือค่าพยากรณ์ ณ เวลา $t + m$ โดยที่ $m = 1$ แทนปี พ.ศ. 2560 ซึ่ง $\alpha = 0.65180$, $\gamma = 0.99944$, $\phi = 0.52485$



ภาพที่ 6 กราฟ ACF และ PACF ของความคลาดเคลื่อนจากการพยากรณ์
วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแตรม

ที่มา: Goganutapon, K. (2020). *A comparative prediction accuracy of birth number of the population Pathum Thani Province*. n.p. [8]

ผลการเปรียบเทียบความแม่นยำของค่าพยากรณ์

จากการใช้ตัวแบบพยากรณ์โดยวิธีบอซ-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลต์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉก ในสมการที่ (7) ถึง สมการที่ (10) สำหรับการพยากรณ์ข้อมูลชุดที่ 2 คืออนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี ตั้งแต่ปี พ.ศ. 2560 ถึง ปี พ.ศ. 2562 จำนวน 3 ค่า นำมาคำนวณค่า MAPE และค่า RMSE แสดงดังตารางที่ 2 พบว่าวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ มีค่า MAPE และค่า RMSE ต่ำที่สุด จากภาพที่ 1 ช่วงปี พ.ศ. 2557 ถึง ปี พ.ศ. 2559 จำนวนการเกิดของประชากร จังหวัดปทุมธานี มีแนวโน้มลดลง เพื่อความแม่นยำของตัวแบบพยากรณ์ นำข้อมูลอนุกรมเวลาจำนวนการเกิดของประชากร จังหวัดปทุมธานี ตั้งแต่ปี พ.ศ. 2557 ถึง ปี พ.ศ. 2562 จำนวน 6 ค่า นำมาคำนวณค่า MAPE และค่า RMSE แสดงดังตารางที่ 3 พบว่าวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ มีค่า MAPE และค่า RMSE ต่ำที่สุด

ตารางที่ 2 ค่าจริงและค่าพยากรณ์ จำนวนการเกิดของประชากร จังหวัดปทุมธานี (คน) ตั้งแต่ปี พ.ศ. 2560 ถึง ปี พ.ศ. 2562 ค่า MAPE และค่า RMSE

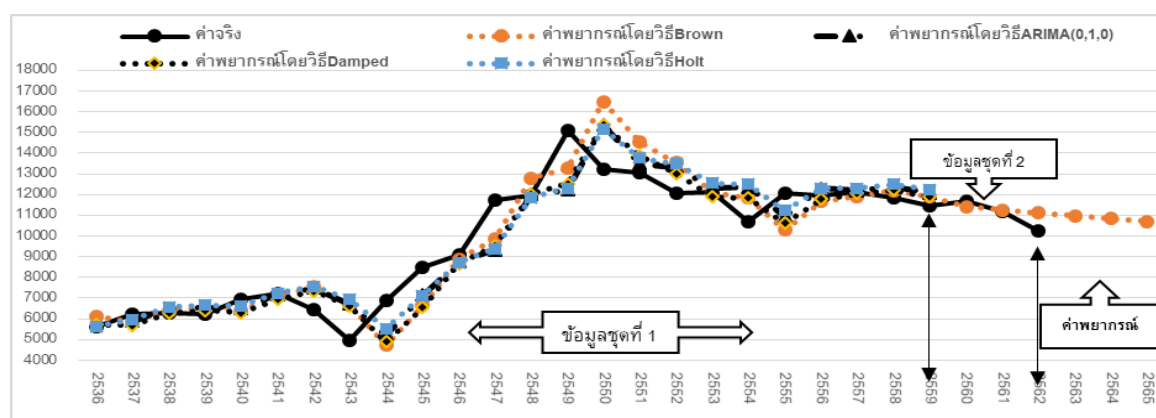
ช่วงเวลา	จำนวนการเกิด ของประชากร	วิธีการพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี			
		ARIMA (0,1,0)	Holt	Brown	Damped Trend
ปี พ.ศ. 2560	11,690	11,736	12,218	11,410	11,481
ปี พ.ศ. 2561	11,183	11,945	12,849	11,266	11,390
ปี พ.ศ. 2562	10,245	11,438	13,479	11,122	11,343
MAPE		6.28	16.99	<u>3.90</u>	4.79
RMSE		11708.08	12858.80	<u>11266.81</u>	11404.89

ตารางที่ 3 ค่าจริงและค่าพยากรณ์ จำนวนการเกิดของประชากร จังหวัดปทุมธานี (คน) ตั้งแต่ปี พ.ศ. 2557 ถึง ปี พ.ศ. 2562 ค่า MAPE และค่า RMSE

ช่วงเวลา	จำนวนการเกิด ของประชากร	วิธีการพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี			
		ARIMA (0,1,0)	Holt	Brown	Damped Trend
ปี พ.ศ. 2557	12130	12184	12311	11907	12052
ปี พ.ศ. 2558	11823	12385	12492	12189	12221
ปี พ.ศ. 2559	11481	12078	12234	11874	11887
ปี พ.ศ. 2560	11690	11736	12218	11410	11,481
ปี พ.ศ. 2561	11183	11945	12849	11266	11,390
ปี พ.ศ. 2562	10245	11438	13479	11122	11,343
MAPE		4.87	10.78	<u>3.3425</u>	3.65
RMSE		11964.86	12605.05	<u>11634.44</u>	11733.98

สรุปและอภิปรายผล

การวิจัยครั้งนี้ได้นำเสนอวิธีการสร้างและคัดเลือกตัวแบบพยากรณ์ที่มีความแม่นยำสำหรับการพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี โดยใช้ข้อมูลจากเว็บไซต์ของ ระบบสถิติทางการทะเบียน สำนักบริหารการทะเบียน กรมการปกครอง ตั้งแต่ ปี พ.ศ. 2536 ถึงปี พ.ศ. 2562 จำนวน 27 ค่า ข้อมูลแบ่งออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่ปี พ.ศ. 2536 ถึงปี พ.ศ. 2559 จำนวน 24 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วย วิธีบอซ-เจนกินส์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของโฮลท์ วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉก และข้อมูลชุดที่ 2 ตั้งแต่ปี พ.ศ. 2560 ถึง ปี พ.ศ. 2562 จำนวน 3 ค่า สำหรับการเปรียบเทียบความแม่นยำของวิธีการพยากรณ์ที่มีความแม่นยำโดยใช้เกณฑ์ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ที่ต่ำที่สุด ผลการวิจัยพบว่า วิธีการพยากรณ์ที่มีความแม่นยำมากที่สุดคือวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของ บราวน์ ซึ่งมีตัวแบบพยากรณ์เป็น $\hat{Y}_{t+m} = 11664.24334 - 144.10576 \left[(m-1) + \frac{1}{0.56757} \right]$ เมื่อ m แทนจำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า โดยมีค่าเริ่มต้น คือ ปี พ.ศ. 2560 ($m = 1$) เมื่อใช้วิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังด้วยวิธีของบราวน์ พยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี ตั้งแต่ปี พ.ศ. 2563 ถึง ปี พ.ศ. 2565 พบว่า จำนวนการเกิดของประชากร จังหวัดปทุมธานี มีแนวโน้มลดลง และแสดงการเปรียบเทียบผลการพยากรณ์ทั้ง 4 วิธี ดังภาพที่ 7 อย่างไรก็ตามจำนวนการเกิดของประชากร จังหวัดปทุมธานี มีการเปลี่ยนแปลงอยู่ตลอดเวลา อาจเป็นเพราะจำนวนการย้ายเข้าย้ายออกของประชากร ในจังหวัดปทุมธานี ประชากรส่วนใหญ่มีภาวะการมีบุตรยาก หรือปัจจัยอื่น ๆ ที่เกี่ยวข้อง จึงควรพิจารณาตัวแปรเหล่านี้ ได้นำมาสร้างตัวแบบถดถอย (regression model) ซึ่งเป็นทางเลือกในการสร้างตัวแบบพยากรณ์ที่มีความเหมาะสมมากยิ่งขึ้น [9] อีกทั้งถ้ามีข้อมูลจำนวนการเกิดของประชากร จังหวัดปทุมธานี ที่เป็นปัจจุบันมากขึ้น ผู้วิจัยควรนำมาพัฒนาตัวแบบพยากรณ์ให้มีความเป็นปัจจุบัน เพื่อความเหมาะสมของการพยากรณ์ต่อไปในอนาคต



ภาพที่ 7 ความสัมพันธ์ระหว่างค่าจริง กับค่าพยากรณ์ ทั้ง 4 วิธี และค่าพยากรณ์จำนวนการเกิดของประชากร จังหวัดปทุมธานี ช่วงหน้า 3 ปี

ที่มา: Goganutapon, K. (2020). A comparative prediction accuracy of birth number of the population Pathum Thani Province. n.p. [8]

เอกสารอ้างอิง

- [1] UNFPA. (2018). The power of choice. In *The State of World Population 2018*. New York: Author. p. 12.
- [2] Department of Health, Bureau of Reproductive Health. (2019). *Annual Report 2019*. Nonthaburi: Department of Health. p. 11.
- [3] Caleiro, A. (2010). Forecasting the number of births in Portugal. In *Joint Eurostat/UNECE Work Session on Demographic Projections*. Conference of European statisticians (28-30, April 2010, Lisbon, Portugal). pp. 1-9.
- [4] The Bureau of registration administration. (2020). *Number of births*. Pathum Thani: Department of Provincial Administration. p. 1.
- [5] Bowerman, B. L., & O'Connell, R. T. (1993). *Forecasting and time series: an applied approach*. 3rd ed. California: Duxbury Press. pp. 521-565.
- [6] Ket-iam, S. (2005). *Forecasting Technique*. 2nd ed. Songkhla: Thaksin University. p. 328
- [7] Manmin, M. (2006). *Time Series and Forecasting*. Bangkok: Fourprinting. p. 448.
- [8] Goganutapon, K. (2020). *A comparative prediction accuracy of birth number of the population Pathum Thani Province*. n.p.
- [9] Montgomery, D. C., Peck, E. A. & Vining, G. G. (2006). *Introduction to Linear Regression Analysis*. 4th ed. New York: John Wiley Sons. p. 535