

รายงานนิสิตชั้นปีที่ 4

รายงานนี้เป็นส่วนหนึ่งของรายวิชา 2104585 INDUSTRIAL FORECASTING

การพยากรณ์อัตราการฆ่าตัวตายรายปีของประชากรในประเทศไทย (FORECASTING OF YEARLY SUICIDE RATE IN THAILAND)

จัดทำโดย

นายกัณตณัฐ	ไชยสุขเจริญกุล	รหัสประจำตัว	6130032821
นายธีรภัทร์	ลิ้มปชโยพาส	รหัสประจำตัว	6130260521
นายณพนธ์	มังคลาด	รหัสประจำตัว	6130274321

เสนอ

อ. ดร.อมรศิริ วิลาสเดชาพันธ์
(Amonsiri Vilasdachanot, Ph.D.)

ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2564

บทคัดย่อ

ในปัจจุบัน ประเทศไทยถือได้ว่าเป็นประเทศที่มีอัตราการฆ่าตัวตายสูงสุดในภูมิภาคเอเชียตะวันออกเฉียงใต้ ซึ่งประมาณได้ว่าในทุก ๆ 10 นาที จะมีคนที่พยายามฆ่าตัวตายจำนวนหนึ่งคน โดยที่อัตราการฆ่าตัวตายของประเทศไทยมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่องในทุก ๆ ปี โดยเฉพาะในช่วงสถานการณ์การแพร่ระบาดของโรค COVID-19 ที่ส่งผลให้มีอัตราการฆ่าตัวตายเพิ่มสูงขึ้นอย่างเห็นได้ชัด โดยที่มีจำนวน 5,085 คน ในปี 2563 เพิ่มขึ้นกว่า 11% เมื่อเปรียบเทียบกับในปี 2562 ที่มีจำนวน 4,581 คน โดยพบว่าในเพศชายมีอัตราการฆ่าตัวตายที่สูงกว่าผู้หญิงถึง 4 เท่า ตามรายงานของกระทรวงสาธารณสุข ซึ่งมีสาเหตุมาจากความสัมพันธ์กับคนใกล้ชิด สภาวะซึมเศร้า และผลกระทบจากปัญหาเศรษฐกิจ ที่ส่งผลให้เกิดการขาดรายได้ และการตกงาน

ทางคณะผู้จัดทำจึงเล็งเห็นถึงปัญหาอัตราการฆ่าตัวตายที่มีแนวโน้มเพิ่มสูงขึ้นเรื่อย ๆ ดังนั้นทางคณะผู้จัดทำจึงมีจุดประสงค์ที่จะพยากรณ์อัตราการฆ่าตัวตายเป็นรายปีของประชากรในประเทศไทย โดยที่ทำการศึกษาผ่านปัจจัยที่คาดว่าจะมีความเกี่ยวข้อง ไม่ว่าจะเป็น ปัจจัยทางด้านประชากร ปัจจัยทางด้านเศรษฐกิจ ปัจจัยทางการพัฒนา ปัจจัยทางด้านแรงงาน และอื่น ๆ โดยใช้วิธีการพยากรณ์วิธีต่าง ๆ เพื่อทำการวิเคราะห์ และเปรียบเทียบ เพื่อที่จะหาวิธีที่ทำให้ได้ผลแม่นยำ และถูกต้องมากที่สุด ไม่ว่าจะเป็นวิธี Time Series method, Regression model, BOX-JENKINS (ARIMA) และ Judgmental Forecasting

โดยจากผลของการดำเนินงาน พบว่าค่าพยากรณ์อัตราการฆ่าตัวตายเป็นรายปีของประชากรในประเทศไทย จากวิธี Multiple regression model มีค่าความแม่นยำ โดยวัดจากค่า Adjust R-squared อยู่ที่ 84.35% โดยที่มีค่าความคลาดเคลื่อน โดยดูจากค่า MAPE อยู่ที่ 5.04% ซึ่งมีความแม่นยำ และประสิทธิภาพดีที่สุด เมื่อเทียบกับวิธีพยากรณ์แบบอื่น ๆ ดังนั้นทางคณะผู้จัดทำจึงมีความตั้งใจจะนำแบบจำลองพยากรณ์ดังกล่าว นำไปใช้กับหน่วยงานที่มีความเกี่ยวข้อง ไม่ว่าจะเป็นกรมสุขภาพจิต หรือกระทรวงสาธารณสุข เพื่อที่จะช่วยลดความรุนแรงของปัญหาอัตราการฆ่าตัวตาย และสามารถวิเคราะห์ปัจจัยต่าง ๆ ที่มีความเกี่ยวข้องได้อย่างทัน่วงที เพื่อที่จะสามารถออกมาตรการ หรือนโยบายต่าง ๆ เพื่อช่วยเหลือกลุ่มคนที่มีความเสี่ยงได้

สารบัญ

บทคัดย่อ	1
สารบัญ	2
บทที่ 1 บทนำ	3
1.1 ที่มาและความสำคัญ	
1.2 วัตถุประสงค์	
1.3 ขอบเขตรายงาน	
1.4 ประโยชน์ของรายงาน	
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง	5
บทที่ 3 วิธีการดำเนินงาน	9
บทที่ 4 ผลการดำเนินงาน และวิเคราะห์ผลการดำเนินงาน	19
บทที่ 5 สรุปผลการดำเนินงาน	43
เอกสารอ้างอิง	44

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

การฆ่าตัวตาย เป็นการกระทำที่ทำให้ตนเองถึงแก่ความตายอย่างตั้งใจ การฆ่าตัวตายเป็นการเกิดจากภาวะซึมเศร้า ซึ่งเกิดจากความผิดปกติทางจิต เช่น โรคซึมเศร้า โรคอารมณ์สองขั้ว หรือการใช้สารเสพติด ปัจจัยที่ทำให้เกิดความเครียด เช่น ความลำบากทางการเงิน หรือปัญหากับความสัมพันธ์ระหว่างบุคคลก็มีส่วนเช่นกัน

ในปัจจุบัน การฆ่าตัวตายถือได้ว่าเป็นพฤติกรรมที่ทางด้านการสาธารณสุข และการแพทย์ทั่วโลกกำลังให้ความสำคัญ หลายครั้งเมื่อมีข่าวการฆ่าตัวตาย ผู้คนมักให้ความสำคัญกับสาเหตุของปัญหาในการฆ่าตัวตาย และสื่อต่าง ๆ มักให้ความสำคัญกับเรื่องนี้ จากข้อมูลของทางองค์การอนามัยโลก หรือ WHO ให้คำอธิบายไว้ว่า ผู้ที่มีความคิดที่อยากจะฆ่าตัวตาย เริ่มต้นมักมีการพยายามฆ่าตัวตายมากกว่า 20 ครั้ง ถึงจะสำเร็จ ดังนั้นผู้เชี่ยวชาญของทางองค์การต่าง ๆ นั้นจึงคิดว่า วิธีต่าง ๆ ที่สามารถระงับเหตุการณ์ฆ่าตัวตายนั้ก่อนที่จะสำเร็จ โดยการรู้สาเหตุของการฆ่าตัวตายนั้ ๆ ก่อน เราก็สามารถที่จะป้องกันการฆ่าตัวตายดังกล่าวได้ อย่างไรก็ตาม การพยายามฆ่าตัวตาย และการฆ่าตัวตายนั้จัดเป็นปัญหาที่เกิดจากสุขภาพจิต ที่หลาย ๆ หน่วยงานเฝ้าระวัง และจับตามองเป็นพิเศษ และได้มีการติดตามเก็บผลลัพธ์มาเป็นระยะเวลานาน ซึ่งอัตราการฆ่าตัวตายนั้ยังถูกกำหนดให้เป็นหนึ่งในตัวชี้วัดที่มีการติดตามเฝ้าระวังโดยหน่วยงานสาธารณสุขทั่วประเทศ

กรมสุขภาพจิต ในฐานะหน่วยงานหลักด้านสุขภาพจิต และจิตเวชของประเทศไทย ได้เฝ้าระวังป้องกันปัญหาการฆ่าตัวตายต่อเนื่องมาเป็นระยะเวลานาน อัตราฆ่าตัวตายจากรายงานการเฝ้าระวังปัญหาในปี 2549 เป็นปีที่อัตราฆ่าตัวตายของประเทศลดต่ำสุดคือ 5.77 ต่อประชากรแสนคน หลังจากนั้นอัตราฆ่าตัวตายอยู่ระดับ 5.97-6.2 มาต่อเนื่อง โดยที่ขึ้นไปสูงสุดในปี 2558 โดยมีอัตราอยู่ที่ 6.47 และข้อมูลล่าสุดคือในปี 2561 อยู่ที่ 6.32 ต่อประชากรแสนคน และในช่วงของสถานการณ์การแพร่ระบาดของ COVID-19 ทำให้กราฟของอัตราการฆ่าตัวตายของประชากรในประเทศไทยนั้นแนวโน้มที่จะเพิ่มสูงขึ้นเรื่อย ๆ

ประเทศไทยถือได้ว่ามีอัตราการฆ่าตัวตายสูงที่สุดในทวีปเอเชียตะวันออกเฉียงใต้ สามารถประมาณได้ว่ามีคนหนึ่งคนพยายามฆ่าตัวตายในทุก ๆ 10 นาที โดยเฉพาะอย่างยิ่งในช่วงสถานการณ์ COVID-19 อัตราการฆ่าตัวตายได้เพิ่มขึ้นจากในปี 2562 เทียบกับสิ้นปี 2563 จากจำนวนผู้เสียชีวิตจาก 4,581 เป็น 5,085 คน นอกจากนี้ในช่วงที่ประกาศล็อกดาวน์ครั้งแรก สายด่วนปรึกษาปัญหาสุขภาพจิตยังได้รับจำนวนสายโทรศัพท์เพิ่มขึ้นจากเดิม 2 เท่า เมื่อเทียบกับในช่วงสถานการณ์ปกติ ดังนั้น หากเราสามารถติดตามอัตราการฆ่าตัวตาย และรู้สาเหตุ หรือปัจจัยที่เกี่ยวข้องกับที่เกี่ยวกับการตัดสินใจฆ่าตัวตายได้ เราก็สามารถที่จะหาวิธีป้องกันที่มีประสิทธิภาพที่จะช่วยลดอัตราการฆ่าตัวตายในประเทศไทยให้น้อยลงกว่าเดิม

1.2 วัตถุประสงค์ของโครงการ

1. เพื่อสร้างแบบจำลองพยากรณ์อัตราการฆ่าตัวตายรายปีของประชากรในประเทศไทยให้มีความถูกต้องและแม่นยำ เพื่อใช้พยากรณ์สำหรับแนวโน้มของอัตราการฆ่าตัวตายในอนาคต
2. เพื่อศึกษาหาความสัมพันธ์ของปัจจัยต่าง ๆ ที่ได้เก็บรวบรวมมา ที่อาจส่งผลกระทบต่อแนวโน้ม หรือการเปลี่ยนแปลงของอัตราการฆ่าตัวตายของประชากรในประเทศไทย
3. เพื่อเปรียบเทียบผลที่ได้จากแบบจำลองพยากรณ์กับข้อมูลทางกรมสุขภาพจิตได้คาดการณ์เอาไว้

1.3 ขอบเขตโครงการ

1. ข้อมูลที่ใช้ศึกษา และสร้างแบบจำลองพยากรณ์ เป็นข้อมูลที่ได้จากรายงานจำนวนการฆ่าตัวตายของประเทศไทย จากกรมสุขภาพจิต กระทรวงสาธารณสุข เป็นข้อมูลรายปีตั้งแต่ปี 2540 ถึง 2563 ซึ่งคิดเป็นจำนวนข้อมูลที่สนใจทั้งหมด 24 ข้อมูล
2. มีการศึกษาปัจจัยต่าง ๆ ที่อาจจะมีความเกี่ยวข้อง โดยแบ่งออกเป็นทั้งหมด 5 กลุ่ม ได้แก่ หมวดประชากร หมวดเศรษฐกิจ หมวดการพัฒนาภายในประเทศ หมวดแรงงาน และอื่น ๆ
3. พัฒนาแบบจำลองพยากรณ์ในมุมมองของกรมสุขภาพจิต กระทรวงสาธารณสุข หรือรัฐบาล เพื่อใช้ในการออกนโยบายต่าง ๆ เพื่อแก้ไข และป้องกันปัญหาการฆ่าตัวตายในประเทศไทย

1.4 ประโยชน์ของโครงการ

1. แบบจำลองพยากรณ์อัตราการฆ่าตัวตายรายปีของประชากรในประเทศไทย
2. สามารถพยากรณ์อัตราการฆ่าตัวตายรายปีโดยพิจารณาจากปัจจัยต่าง ๆ ได้แม่นยำมากยิ่งขึ้น
3. สามารถนำไปประยุกต์ใช้กับองค์กรต่าง ๆ ที่มีส่วนเกี่ยวข้องได้ โดยเฉพาะกรมสุขภาพจิต กระทรวงสาธารณสุข ในการนำเอาผลลัพธ์ของการศึกษาที่ได้ ไปพัฒนาในส่วนของการช่วยสนับสนุนสุขภาพจิต หรือแบบทดสอบภาวะซึมเศร้า ให้ครอบคลุม และมีประสิทธิภาพมากยิ่งขึ้น

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 องค์ประกอบของอนุกรมเวลา (Component of Time-series)

อนุกรมเวลาเป็นข้อมูลที่มีการเก็บรวบรวมข้อมูลเป็นช่วงระยะเวลา และมีลักษณะเป็นข้อมูลที่ต่อเนื่อง เช่น ราคาหุ้นที่มีการเปลี่ยนแปลงตลอดเวลา อาจจะเป็นรายชั่วโมง รายวัน หรือรายปี เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงตลอดเวลา ดังนั้นการพยากรณ์เพื่อให้รู้ผลในอนาคตจึงเป็นขั้นตอนที่สำคัญเพื่อใช้ในการวางแผนตัดสินใจในการทำงานต่าง ๆ ซึ่งองค์ประกอบของอนุกรมเวลามีดังนี้

2.1.1 แนวโน้ม (Trend)

เป็นการเปลี่ยนแปลงข้อมูลมีลักษณะราบเรียบ แนวโน้มอาจมีลักษณะเป็นเส้นตรง หรือเส้นโค้งในทิศทางเพิ่มขึ้น หรือลดลง ค่าแนวโน้มของข้อมูลเป็นการเคลื่อนไหวในช่วงระยะเวลาที่ค่อนข้างนานพอสมควร

2.1.2 ความแปรผันตามฤดูกาล (Seasonal)

เป็นการเปลี่ยนแปลงข้อมูลที่มีลักษณะการเพิ่มขึ้น หรือลดลงในลักษณะเดียวกันของรอบระยะเวลาหนึ่งที่แน่นอน เรียกว่า การเปลี่ยนแปลงตามฤดูกาล หน่วยของระยะเวลาสำหรับข้อมูลอาจเป็นรายชั่วโมง รายวัน รายสัปดาห์ รายเดือน รายไตรมาส สำหรับข้อมูลรายปีไม่มีการแปรผันตามฤดูกาล การเปลี่ยนแปลงตามฤดูกาลนั้นกำหนดระยะเวลาการเกิดซ้ำในรอบหนึ่ง ๆ ได้ค่อนข้างแน่นอน

2.1.3 ความผันแปรตามวัฏจักร (Cyclical)

เป็นข้อมูลที่มีลักษณะผันผวนในช่วงระยะเวลาหนึ่ง มีลักษณะที่ขึ้นหรือลงแบบไม่มีคาบเวลาที่ตายตัว โดยปกติแล้วจะใช้เวลาหลายปีถึงจะเกิดวัฏจักร อาจเรียกได้ว่ากราฟของข้อมูลเป็นแบบ wavelike pattern

2.1.4 ความผันแปรเนื่องจากเหตุการณ์ผิดปกติ (Irregular)

หรือเรียกอีกอย่างหนึ่งว่าเป็นการเปลี่ยนแปลงแบบสุ่ม เป็นลักษณะที่ไม่สามารถคาดการณ์ได้ว่ากราฟจะเป็นรูปแบบใด มีความผันผวนที่คาดการณ์ไม่ได้

2.2 ทฤษฎีที่เกี่ยวข้องกับวิธีพยากรณ์อนุกรมเวลา (Time-series method)

2.2.1 Naïve with trend method

เป็นการพยากรณ์ที่เหมาะสมกับชุดข้อมูลปริมาณไม่มาก โดยวิธีพยากรณ์นั้นจะให้ความสำคัญกับข้อมูลล่าสุด กล่าวคือจะอิงตามข้อมูลปัจจุบันมากกว่าอดีต ซึ่งสามารถเปลี่ยนแปลงได้อย่างรวดเร็ว ส่วนใหญ่จะใช้เป็น model ในการพยากรณ์เพื่อเปรียบเทียบกับวิธีการพยากรณ์อื่น ๆ

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1})$$

โดยที่:

Y_t = ค่าของข้อมูลในช่วงเวลา t ใด ๆ

2.2.2 Simple exponential smoothing

เป็นการพยากรณ์ที่ใช้ในการหาค่าเฉลี่ยแบบถ่วงน้ำหนักให้ออกมาเป็นรูปสมการที่ใช้ในการคำนวณ ซึ่งจะถ่วงน้ำหนักถ่วงค่า α (α) ซึ่งจะมีค่าระหว่าง 0-1 ซึ่งถ้าค่า α มีค่าเข้าใกล้ 1 ก็แปลว่าการถ่วงน้ำหนักนั้นให้ความสำคัญกับข้อมูลปัจจุบันมากกว่า และถ้าค่า α มีค่าเข้าใกล้ 0 ค่าพยากรณ์ที่ได้จะมีความใกล้เคียงกับค่าในอดีตมากกว่าปัจจุบัน

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) Y_t$$

โดยที่:

Y_t = ค่าของข้อมูลในช่วงเวลา t ใด ๆ

α = ค่าสัมประสิทธิ์ที่ใช้ในการถ่วงน้ำหนักในสมการคำนวณ

2.2.3 Holt's linear method

เป็นวิธีการพยากรณ์ที่มีลักษณะคล้าย Simple exponential smoothing แต่ว่าจะมีการปรับปรุงเพื่อนำไปใช้สำหรับข้อมูลที่มีลักษณะเป็น trend

$$\begin{aligned} L_t &= \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \alpha)T_{t-1} \\ \hat{Y}_{t+p} &= L_t + pT_t \end{aligned}$$

โดยที่:

L_t = level estimate ในช่วงเวลา t ใด ๆ

T_t = trend estimate ในช่วงเวลา t ใด ๆ

\hat{Y}_{t+p} = ค่าของข้อมูลในช่วงเวลา t+p

α = ค่าสัมประสิทธิ์ที่ใช้ในการถ่วงน้ำหนักในสมการคำนวณ

β = ค่าสัมประสิทธิ์ที่ใช้ในการถ่วงน้ำหนักในสมการคำนวณ

2.2.4 ARIMA (Autoregressive integrated moving average)

เป็นวิธีการพยากรณ์ที่ถูกพัฒนาขึ้นเพื่อใช้กับข้อมูลประเภท Time-series คือการพยากรณ์นั้นจะใช้ข้อมูลในอดีต โดยคำนึงถึงรูปแบบต่าง ๆ ของข้อมูล เพื่อกำหนด parameters ในสมการ

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q}$$

โดยที่:

Y_t = ค่าของข้อมูลในช่วงเวลา t ใด ๆ

C = ค่าคงที่

ϕ_p = ค่าสัมประสิทธิ์ของแต่ละพารามิเตอร์ p

ε_t = ค่าความคลาดเคลื่อนในช่วงเวลา t ใด ๆ

ω_q = ค่าสัมประสิทธิ์ของแต่ละพารามิเตอร์ q

2.3 ทฤษฎีที่เกี่ยวข้องกับวิธีพยากรณ์โดยใช้หลักของเหตุและผล (Causal method)

เป็นการศึกษาเกี่ยวกับความสัมพันธ์ของตัวแปร วัตถุประสงค์หลักของการวิเคราะห์สมการถดถอยคือ ต้องการประมาณค่าของตัวแปรตัวหนึ่ง ๆ ซึ่งเรียกว่า ตัวแปรตาม (Dependent variable or response variable) เขียนแทนด้วย Y โดยอาศัยข้อมูลจากตัวแปรอื่น ๆ ซึ่งเรียกว่าตัวแปรอิสระ (Independent variable) เขียนแทนด้วย X ซึ่งถ้าใช้ตัวแปร X เพียงตัวแปรเดียวในการประมาณ Y และความสัมพันธ์ของ Y และ X เป็นเชิงเส้นตรง เรียกว่า การถดถอยเชิงเส้นอย่างง่าย (Simple linear regression) แต่ถ้าถ้าใช้ X มากกว่า 1 ตัว ในการทำนาย Y และความสัมพันธ์ของ X แต่ละตัวกับ Y มีความสัมพันธ์กันเชิงเส้นตรง จะเรียกว่า การถดถอยพหุคูณ (Multiple regression)

2.3.1 Simple linear regression

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

โดยที่:

Y = ค่าของตัวแปรตาม

β_0 = ค่าคงที่ของสมการถดถอยเชิงเส้นอย่างง่าย

β_1 = ค่าสัมประสิทธิ์ของสมการถดถอยเชิงเส้นอย่างง่าย

ε = ค่าความคลาดเคลื่อน

X = ตัวแปรอิสระ

2.3.2 Multiple regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

โดยที่:

Y = ค่าของตัวแปรตาม

β_0 = ค่าคงที่ของสมการถดถอยเชิงเส้นอย่างง่าย

β_k = ค่าสัมประสิทธิ์ของสมการถดถอยเชิงเส้นอย่างง่ายของแต่ละ k ตัว

ε = ค่าความคลาดเคลื่อน

X_k = ตัวแปรอิสระของแต่ละ k ตัว

2.4 การวัดค่าความแม่นยำ (Forecasting error measurements)

โดยวิธีที่ทางผู้จัดทำใช้วัดค่าความแม่นยำ คือ การประเมินผลความคลาดเคลื่อนในการทำนายโดยใช้ Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Mean Percentage Error (MPE), Root Mean Square Error (RMSE) และ Mean Square Error (MSE)

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}$$

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}$$

บทที่ 3

วิธีการดำเนินงาน

3.1 Data collection

ทางผู้จัดทำได้เก็บรวบรวมข้อมูลอัตราการฆ่าตัวตายรายปีของประชากรในประเทศไทยตั้งแต่ปี 2540 ถึง 2560 รวมทั้งสิ้นเป็นจำนวน 24 ข้อมูล จากรายงานจำนวนการฆ่าตัวตายของประเทศไทย ของกรมสุขภาพจิต กระทรวงสาธารณสุข และได้เก็บรวบรวมข้อมูลปัจจัยที่อาจจะมีส่วนเกี่ยวข้อง แบ่งออกเป็น 5 กลุ่ม ได้แก่ ปัจจัยทางด้านประชากร ปัจจัยทางด้านเศรษฐกิจ ปัจจัยทางด้านการพัฒนาภายในประเทศ ปัจจัยทางด้านแรงงาน และอื่น ๆ โดยได้เก็บรวบรวมข้อมูลในช่วงปีเดียวกันเพื่อนำมาวิเคราะห์ข้อมูล

3.1.1 ข้อมูลอัตราการฆ่าตัวตายรายปีของประชากรในประเทศไทย

year	suicide rate/100k	year	suicide rate/100k
2540	6.92	2552	5.97
2541	8.12	2553	5.90
2542	8.59	2554	6.03
2543	8.40	2555	6.20
2544	7.74	2556	6.08
2545	7.84	2557	6.07
2546	7.13	2558	6.47
2547	6.87	2559	6.35
2548	6.34	2560	6.03
2549	5.77	2561	6.32
2550	5.97	2562	6.64
2551	5.98	2563	7.37

Table 3.1: Yearly suicide rate/100k

3.1.2 ปัจจัยที่คาดว่าจะมีความเกี่ยวข้อง

3.1.2.1 ปัจจัยทางด้านประชากร (Population)

No.	Predictor	Explanation	Unit
1	Growth rate	อัตราการเติบโตของประชากรในประเทศ	
2	Urban people	จำนวนประชากรที่อาศัยอยู่ในเมือง	millions
3	Rural people	จำนวนประชากรที่อาศัยอยู่ในชนบท	millions
4	Life expectancy	อายุคาดการณเฉลี่ย	years
5	Population	จำนวนประชากรในประเทศ	millions

Table 3.2: Population predictors

3.1.2.2 ปัจจัยทางด้านเศรษฐกิจ (Economy)

No.	Predictor	Explanation	Unit
1	Inflation rate	อัตราเงินเฟ้อในประเทศ	percentage
2	Manufacturing	มูลค่าของอุตสาหกรรมการผลิตภายในประเทศ	\$billions
3	GDP per capita	ผลิตภัณฑ์มวลรวมภายในประเทศต่อหัว	\$1,000
4	GNI per capita	รายได้เฉลี่ยต่อหัว	\$1,000
5	Number of tourists	จำนวนนักท่องเที่ยวที่เข้ามาภายในประเทศ	millions
6	Trade balance	มูลค่าของงบดุลการค้า	\$billions
7	Imports	มูลค่าของปริมาณสินค้านำเข้าในประเทศ	10 ⁶ THB
8	Exports	มูลค่าของปริมาณสินค้าส่งออกนอกประเทศ	10 ⁶ THB

Table 3.3: Economy predictors

3.1.2.3 ปัจจัยทางการพัฒนาภายในประเทศ (Development)

No.	Predictor	Explanation	Unit
1	Poverty rate	อัตราของปริมาณคนยากจนภายในประเทศไทย	
2	Hungers	จำนวนประชากรที่มีความหิวโหยเทียบกับทั้งหมด	percentage

Table 3.4: Development predictors

3.1.2.4 ปัจจัยทางด้านแรงงาน (Labor force)

No.	Predictor	Explanation	Unit
1	Unemployment rate	อัตราของคนตกงานภายในประเทศ	
2	Youth employment rate	อัตราตงานของกลุ่มวัยรุ่นในประเทศ	
3	Participation rate	อัตราการมีงานทำของกลุ่มวัยรุ่นภายในประเทศ	

Table 3.5: Labor force predictors

3.1.2.5 ปัจจัยอื่น ๆ (Others)

No.	Predictor	Explanation	Unit
1	Health per capita	มูลค่างบค่าใช้สุงภาพต่อหัวของประชากรในประเทศ	\$US
2	Military	งบการทหารภายในประเทศ	\$billions
3	Crime rate/100k	อัตราการเกิดอาชญากรรมภายในประเทศ	per 100,000

Table 3.6: Other predictors

3.2 Data manipulation and cleansing

poverty rate	hungers	unemployment rate	youth rate	participation rate	crime rate/100k	health per capita	military
44.45	16.60	1.51	2.19	59.24	7.87	62.11	3.26
43.90	16.80	4.37	7.58	56.66	8.71	60.90	2.11
47.80	16.70	4.20	8.01	53.11	8.11	63.15	2.06
48.60	16.30	3.60	6.83	52.92	8.17	62.26	1.88
44.15	17.40	3.66	7.88	53.77	7.90	57.29	1.72
39.70	16.40	2.48	5.93	52.39	7.08	69.90	1.81
36.55	15.10	2.18	5.22	51.12	9.97	76.51	1.89
33.40	13.70	2.52	4.63	52.02	6.57	83.50	1.87
30.10	12.00	1.80	4.84	51.75	7.35	91.44	1.98
26.80	10.70	1.50	4.89	48.98	7.12	104.45	2.44
24.50	10.20	1.40	4.60	49.59	6.70	126.85	3.52
22.50	10.70	1.40	4.94	48.99	5.97	151.51	4.47
20	10.70	1.50	3.76	48.62	5.54	152.49	4.80
17.50	10.50	1.00	2.65	46.34	5.44	172.09	4.96
13.40	9.90	0.68	3.03	47.35	4.90	195.97	5.51
13.50	9.40	0.66	2.85	46.67	4.90	206.46	5.49
11.30	9.10	0.72	1.09	43.39	4.25	213.00	5.90
10.40	8.80	0.84	3.19	42.48	3.87	219.30	5.73
7.00	8.60	0.88	3.15	41.55	3.87	214.21	5.72
8.20	8.60	0.99	3.76	41.17	3.47	224.99	5.88
7.60	8.90	1.18	4.41	40.54	3.23	252.08	6.32
8.40	9.30	1.05	4.05	41.45	2.58	275.92	6.88
6.20	8.93	0.98	4.22	40.30	3.09	251.00	7.31
7.99	9.04	1.69	4.23	40.76	2.97	259.66	6.84

Table 3.7: Example of predictors collected

จากข้อมูลปัจจัยที่ทางผู้จัดทำคาดว่าจะมีส่วนเกี่ยวข้องกับอัตราการฆ่าตัวตายที่ผู้จัดทำได้เก็บรวบรวมมา พบว่ามีข้อมูลในบางปีที่อยู่ในช่วงปี 2540 ถึง 2563 ที่เป็นช่วงข้อมูลที่อยู่ในขอบเขตของรายงาน ซึ่งในบางปีไม่มีข้อมูลที่ได้เก็บบันทึกเอาไว้ ดังนั้นทางผู้จัดทำจึงทำการ fill missing values โดยการใช้ค่าเฉลี่ยของ 3 ปีที่ติดกัน เพื่อนำไปใช้เป็นข้อมูลที่หายไปในปี ๆ นั้น เพื่อให้การพยากรณ์มีความแม่นยำมากยิ่งขึ้น

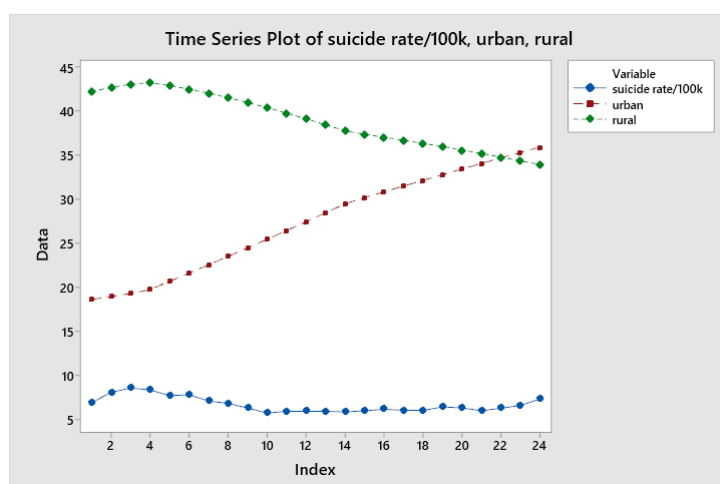


Figure 3.1: Time-series plot of suicide rate/100k, urban and rural

และทางผู้จัดทำก็ได้มีการพิจารณาโดยการตัดตัวแปร Urban people และ Rural people ออกก่อนที่จะนำไปทำแบบจำลองพยากรณ์ เนื่องจากทางผู้จัดทำได้ทำการพิจารณาความสัมพันธ์ของตัวแปรทั้งสองเทียบกับตัวแปรตามที่เราสนใจ ซึ่งก็คืออัตราการฆ่าตัวตายรายปี ซึ่งจาก Time-series plot ที่ได้จาก Minitab จะเห็นได้ว่า ข้อมูลค่อนข้างมีความเป็น trend สัมพันธ์แค่ในชุดข้อมูลของตัวเอง ไม่มีความสัมพันธ์กับตัวแปรตามที่เราสนใจ ดังนั้นเราจึงทำการตัดออก เพื่อป้องกันปัญหา Multicollinearity ในแบบจำลองพยากรณ์

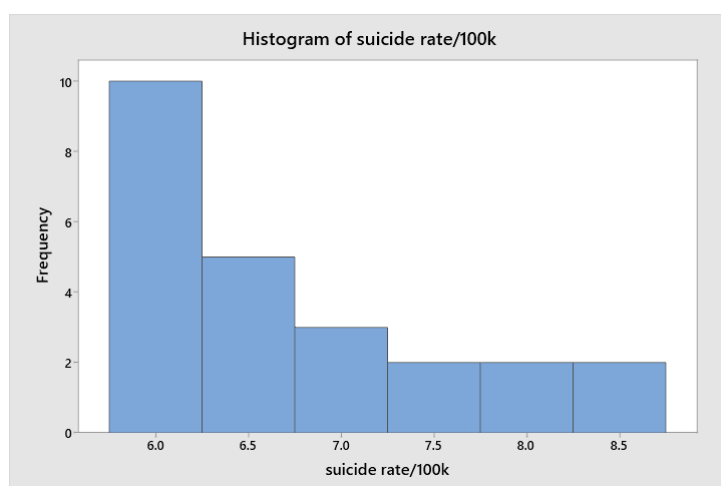


Figure 3.2: Histogram of suicide rate/100k

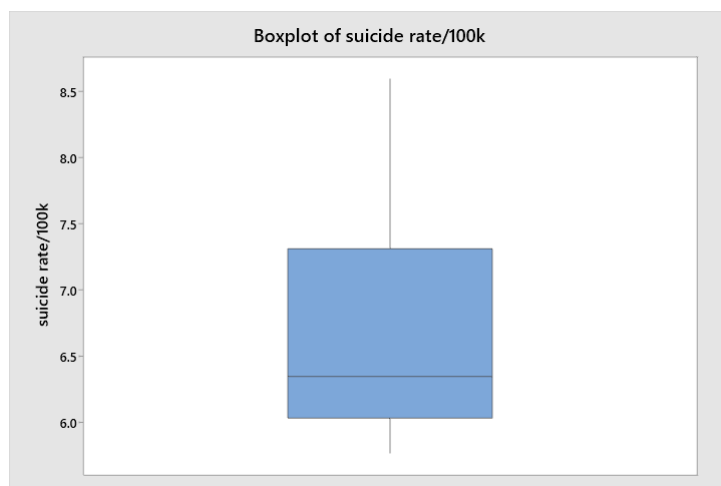
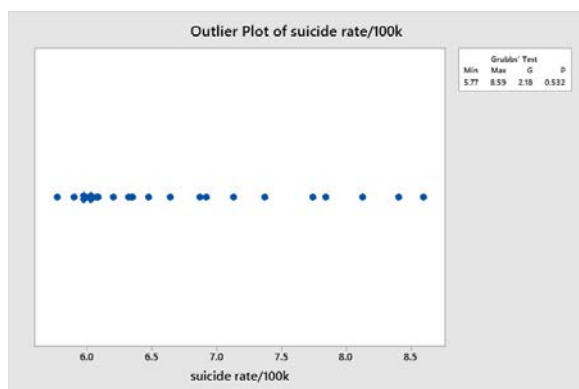


Figure 3.3: Boxplot of suicide rate/100k



Method

Null hypothesis: All data values come from the same normal population
 Alternative hypothesis: Smallest or largest data value is an outlier
 Significance level: $\alpha = 0.05$

Grubbs' Test

Variable	N	Mean	StDev	Min	Max	G	P
suicide rate/100k	24	6.712	0.861	5.770	8.590	2.18	0.532

* NOTE * No outlier at the 5% level of significance

Figure 3.4: Outlier test for suicide rate series

จากผลของ histogram plot ของข้อมูลอัตราการฆ่าตัวตาย ที่ได้จาก Minitab จะเห็นได้ว่าค่าเฉลี่ยของข้อมูลดังกล่าวอยู่ที่ 6.0 เป็นส่วนใหญ่ ซึ่งมี frequency อยู่ที่ประมาณ 10 ตัว และจากผลที่เราได้จากการทำ outlier test ที่ 5% level of significance จะเห็นได้ว่าข้อมูลทั้ง 24 ปี ที่ได้เก็บรวบรวมมา ผลปรากฏว่าไม่พบข้อมูลที่เป็นค่าผิดปกติ หรือ outliers ในชุดข้อมูลดังกล่าว

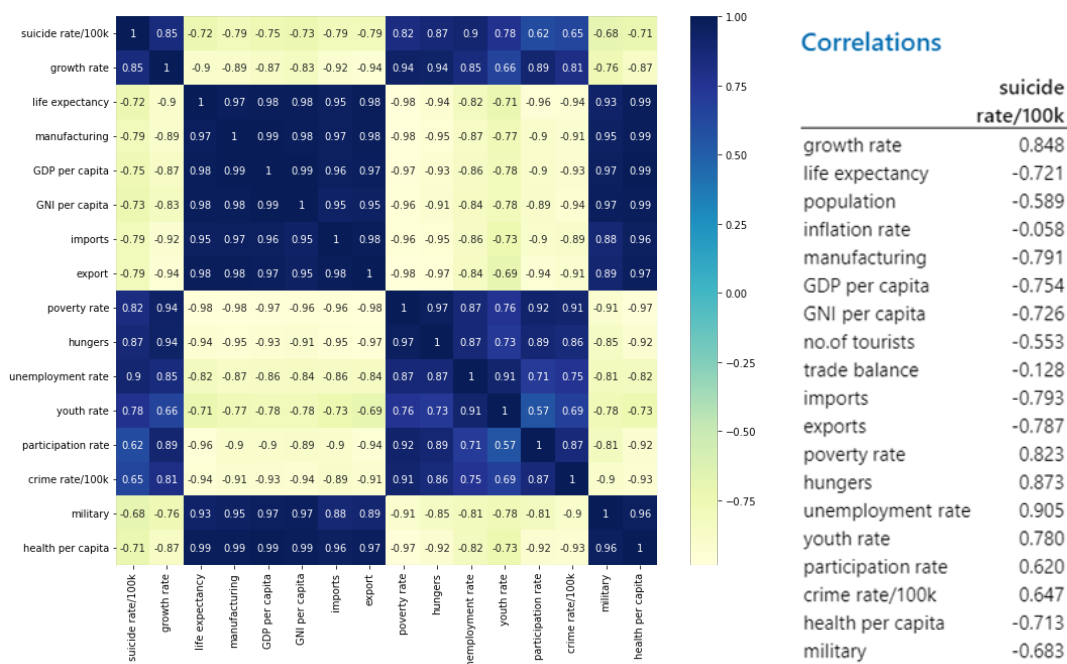


Figure 3.5: Correlation matrix

จากผลของ Correlation matrix ที่เราได้จาก Python จะเห็นได้ว่าข้อมูลของปัจจัยต่าง ๆ ที่ผู้จัดทำได้เก็บมาก่อนข้างที่จะมีความสัมพันธ์ระหว่างกันเอง หรือก็คือมี Multicollinearity เกิดขึ้นภายในชุดข้อมูล ซึ่งผู้จัดทำจะทำการพิจารณา Multicollinearity อย่างละเอียด ในขั้นตอนของการทำ Multiple linear regression model แต่ในเบื้องต้น ผู้จัดทำจะทำการตัดตัวแปรที่มี correlation น้อย ๆ ออกก่อน โดยใช้เกณฑ์คือ ถ้า correlation ของตัวแปรนั้น ๆ เมื่อเทียบกับตัวแปรที่สนใจ มีค่าน้อยกว่า 0.60 ให้ทำการตัดออก ซึ่งจากเกณฑ์ที่ผู้จัดทำได้ตั้งไว้ ทำให้สามารถตัดตัวแปรออกไปทั้งหมด 4 ตัว และคงเหลืออยู่ทั้งหมด 15 ตัว

No.	Predictors removed	Correlation
1	Population	-0.589
2	Inflation rate	-0.058
3	Number of tourists	-0.553
4	Trade balance	-0.128

Table 3.8: Predictors removed

No.	Remaining predictors	Correlation
1	Unemployment rate	0.905
2	Hungers	0.873
3	Growth rate	0.848
4	Poverty rate	0.823
5	Imports	-0.793
6	Manufacturing	-0.791
7	Exports	-0.787
8	Youth rate	0.780
9	GDP per capita	-0.754
10	GNI per capita	-0.726
11	Life Expectancy	-0.721
12	Health per Capita	-0.713
13	Military	-0.680
14	Crime Rate/100k	0.647
15	Participation Rate	0.620

Table 3.9: Remaining predictors

3.3 Examine data patterns

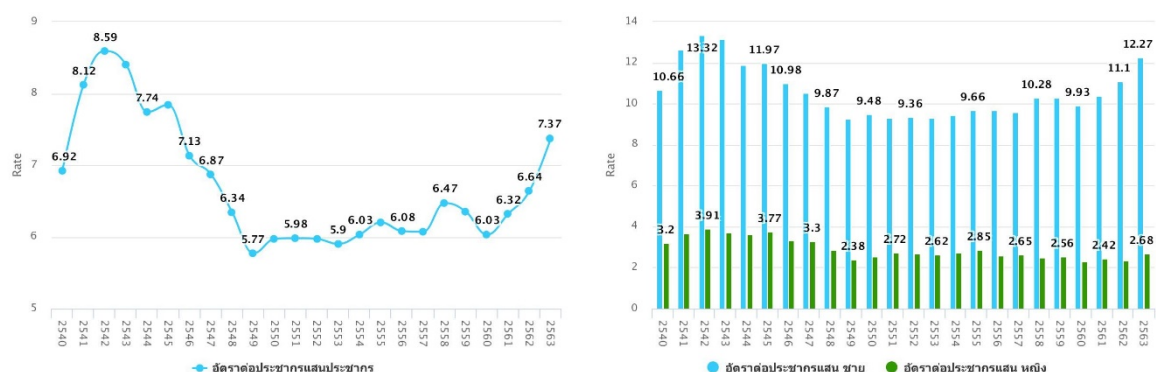


Figure 3.6: Suicide rate plot from Department of Mental Health

จาก suicide rate plot ที่ได้จากรายงานอัตราการฆ่าตัวตาย (รายต่อแสนประชากร) ตั้งแต่ปี 2540 ถึง 2563 จะสามารถสังเกตเห็นได้ว่าข้อมูลมีแนวโน้มของ trend ขึ้นลงในบางช่วง โดยมีอัตราการฆ่าตัวตายต่ำที่สุดในปี 2549 ซึ่งเท่ากับ 5.77 และสูงสุดในปี 2542 ซึ่งมีค่าเท่ากับ 8.59 และจะเห็นว่าในช่วงตั้งแต่ปี 2560 เป็นต้นมา อัตราการฆ่าตัวตายของประชากรในประเทศไทยเพิ่มขึ้นอย่างต่อเนื่อง และก้าวกระโดด และเมื่อเทียบอัตราการฆ่าตัวตายของประชากรชาย และประชากรหญิง เราจะเห็นว่าในประชากรชาย มีอัตราการฆ่าตัวตายสูงกว่าในทุก ๆ ปี อย่างเห็นได้ชัด โดยเฉลี่ยแล้วสูงกว่าในผู้หญิงถึง 4 เท่า

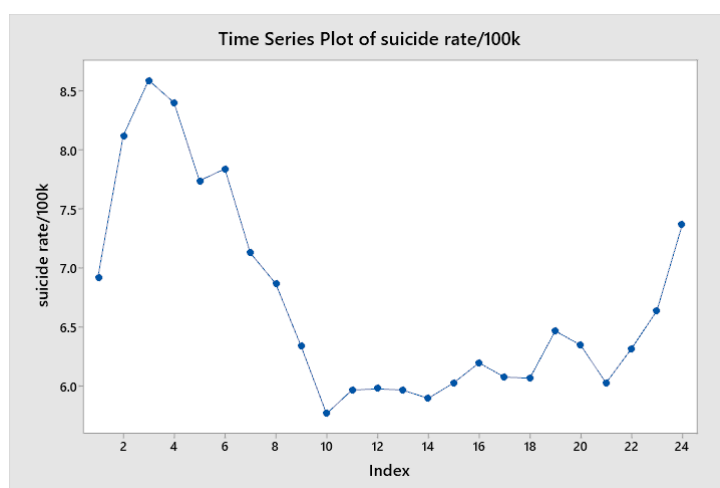


Figure 3.7: Time-series plot of suicide rate/100k

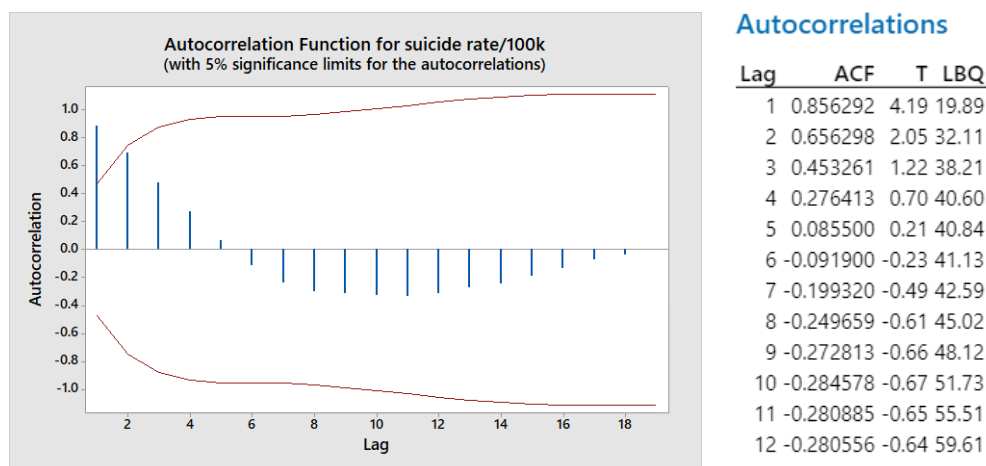


Figure 3.8: Autocorrelation Function for suicide rate/100k

จากผลของ Time-series plot และ Autocorrelation function ที่ได้จาก Minitab จะสังเกตเห็นได้ว่าข้อมูลค่อนข้างมีความสัมพันธ์เป็น trend ที่ลดลงอย่างรวดเร็ว และที่ time lag ที่ 2 จะเห็นได้ว่าค่า ACF ค่อนข้างที่จะติดเส้นสีแดง หรือก็คือ significant ดังนั้นผู้จัดทำจึงได้ทำการวิเคราะห์ค่า ACF ที่ time lag ที่ 2 โดยการทำ hypothesis testing โดยใช้ T-test (Student's T-test) โดยที่ค่า T-observed มีค่าเท่ากับ 2.05 และค่า T-test มีค่าเท่ากับ 2.069 ซึ่งจะเห็นได้ว่าค่า T-observed มีค่าน้อยกว่า T-test เพียงเล็กน้อย ดังนั้นทางผู้จัดทำจึงได้ทำการพิจารณาชุดข้อมูลของ suicide rate series เป็น non-stationary data เนื่องจากว่าเมื่อนำไปใช้ในแบบจำลองพยากรณ์ แล้วได้ค่าความแม่นยำที่สูงกว่าการพิจารณาเป็น stationary data

3.4 Choose forecasting methods

หลังจากที่ผู้จัดทำได้ทำการวิเคราะห์ และสรุปผล pattern ของข้อมูล ที่จะพิจารณา pattern ของข้อมูลเป็น non-stationary data ดังนั้นทางผู้จัดทำจึงได้เลือกวิธี forecasting methods เพื่อให้ครอบคลุม และมีความสอดคล้องกับ pattern ของข้อมูลที่เรากำลังพิจารณา ซึ่งผู้จัดทำได้ใช้ทั้งวิธี Time-series, Causal method or Regression model, ARIMA และมีการพิจารณาในส่วนของการ Judgmental ประกอบด้วย

3.4.1 Time-series method

ในส่วนของการ Time-series method ผู้จัดทำได้พิจารณาส่วนของ trend pattern สำหรับชุดข้อมูลของอัตราการฆ่าตัวตาย ดังนั้นผู้จัดทำจึงได้เลือก forecasting methods ทั้งหมด 3 วิธี ได้แก่ Naïve with trend method, Simple exponential smoothing method และ Holt's linear method

3.4.2 Causal method or Regression model

เนื่องจากข้อมูลอัตราการฆ่าตัวตายเป็นข้อมูล time-series data ดังนั้นผู้จัดทำจึงได้ทำการวิเคราะห์วิธีพยากรณ์แบบ Causal method โดยการพิจารณาโดยใช้วิธี Regression with time-series แบ่งออกได้เป็น 2 วิธี ได้แก่ Simple และ Multiple linear regression model

3.4.3 The BOX-JENKINS (ARIMA) methodology

ในส่วนของ ARIMA method ที่ผู้จัดทำจะกล่าวถึงรายละเอียดในบทต่อไป ผู้จัดทำได้ทำการพิจารณาแบบจำลองพยากรณ์ของวิธี ARIMA โดยอ้างอิงจาก pattern ของข้อมูลอัตราการฆ่าตัวตายรายปีที่ได้ทำการวิเคราะห์ ออกเป็นทั้งหมด 3 แบบจำลอง ได้แก่ ARIMA(0,1,0), ARIMA(1,0,0) และ ARIMA(2,0,0)

3.5 Splitting data for overfitting check

ทางผู้จัดทำได้ทำการแบ่งข้อมูลของอัตราการฆ่าตัวตายรายปี และ remaining predictors ที่ส่งต่อผลต่อตัวแปรที่เราสนใจดังกล่าว ในช่วงของข้อมูลเดียวกัน เพื่อที่จะทำการพิจารณา overfitting check ของ forecasting method ที่เลือก โดยผู้จัดทำได้ทำการแบ่งข้อมูลออกเป็น 80:20 ซึ่งจะได้จำนวนข้อมูลที่เป็น train set ทั้งหมด 20 ข้อมูล (2540-2559) และสำหรับ test set เป็นจำนวน 4 ข้อมูล (2560-2563)

Train set				Test set	
year	suicide rate/100k	year	suicide rate/100k	year	suicide rate/100k
2540	6.92	2552	6.92	2560	6.03
2541	8.12	2553	8.12	2561	6.32
2542	8.59	2554	8.59	2562	6.64
2543	8.40	2555	8.40	2563	7.73
2544	7.74	2556	7.74		
2545	7.84	2557	7.84		
2546	7.13	2558	7.13		
2547	6.87	2559	6.87		
2548	6.34				
2549	5.77				
2550	5.97				
2551	5.98				

Table 3.9: Train and test set from splitting data

บทที่ 4

ผลการดำเนินงาน และวิเคราะห์ผลการดำเนินงาน

4.1 Time-series method

4.1.1 Naïve with trend method

$$\hat{Y}_{t+1} = \hat{Y}_t + (\hat{Y}_t - \hat{Y}_{t-1})$$

จากสมการ Naïve with trend method ที่กล่าวมา ทางผู้จัดทำได้ใช้สมการพยากรณ์ดังกล่าวกับข้อมูลอัตราการฆ่าตัวตายรายปีตั้งแต่ปี 2540 ถึง 2559 ที่เป็นตัว train set และได้ใช้สมการพยากรณ์ดังกล่าวกับตัว test set เพื่อเปรียบเทียบค่า forecast errors ระหว่าง train set กับ test เพื่อพิจารณา overfitting ของสมการพยากรณ์ Naïve with trend method ดังกล่าว โดยใช้ Microsoft Excel

year	suicide rate/100k	\hat{Y}_{t-1}	e_t	e_t^2	abs	abs/ \hat{Y}_t	e_t/\hat{Y}_t
2540	6.92						
2541	8.12						
2542	8.59	9.32	-0.73	0.5329	0.7300	0.0850	-0.0850
2543	8.40	9.06	-0.66	0.4356	0.6600	0.0786	-0.0786
2544	7.74	8.21	-0.47	0.2209	0.4700	0.0607	-0.0607
2545	7.84	7.08	0.76	0.5776	0.7600	0.0969	0.0969
2546	7.13	7.94	-0.81	0.6561	0.8100	0.1136	-0.1136
2547	6.87	6.42	0.45	0.2025	0.4500	0.0655	0.0655
2548	6.34	6.61	-0.27	0.0729	0.2700	0.0426	-0.0426
2549	5.77	5.81	-0.04	0.0016	0.0400	0.0069	-0.0069
2550	5.97	5.20	0.77	0.5929	0.7700	0.1290	0.1290
2551	5.98	6.17	-0.19	0.0361	0.1900	0.0318	-0.0318
2552	5.97	5.99	-0.02	0.0004	0.0200	0.0034	-0.0034
2553	5.90	5.96	-0.06	0.0036	0.0600	0.0102	-0.0102
2554	6.03	5.83	0.20	0.0400	0.2000	0.0332	0.0332
2555	6.20	6.16	0.04	0.0016	0.0400	0.0065	0.0065
2556	6.08	6.37	-0.29	0.0841	0.2900	0.0477	-0.0477
2557	6.07	5.96	0.11	0.0121	0.1100	0.0181	0.0181
2558	6.47	6.06	0.41	0.1681	0.4100	0.0634	0.0634
2559	6.35	6.87	-0.52	0.2704	0.5200	0.0819	-0.0819

Table 4.1: Train set of Naïve with trend method

year	suicide rate/100k	\hat{Y}_t	e_t	e_t^2	abs	abs/ \hat{Y}_t	e_t/\hat{Y}_t
2560	6.03	6.23	-0.20	0.0400	0.2000	0.0332	-0.0332
2561	6.32	5.71	0.61	0.3721	0.6100	0.0965	0.0965
2562	6.64	6.61	0.03	0.0009	0.0300	0.0045	0.0045
2563	7.37	6.96	0.41	0.1681	0.4100	0.0556	0.0556

Table 4.2: Test set of Naïve with trend method

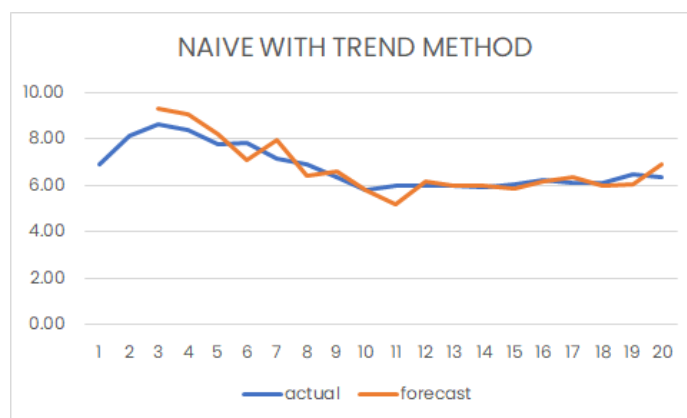


Figure 4.1: Train set plot of Naïve with trend method

Forecast error	Train set	Test set
MAD	0.3778	0.3125
MSE	0.2172	0.1453
RMSE	0.4460	0.3811
MAPE	5.4156%	4.7459%
MPE	-0.8319%	3.0875%

Table 4.3: Forecast errors of train versus test set

จากผลของการเปรียบเทียบ forecast error ของตัว train set กับ test set จะเห็นได้ว่าค่าแนวโน้มของค่า forecast error แต่ละตัว ของตัว test set มีค่าน้อยกว่าตัว train set ทำให้สามารถสรุปได้ว่า สมการพยากรณ์ของวิธี Naïve with trend method ที่นำมาใช้ นั้นไม่ปรากฏผลของ overfitting

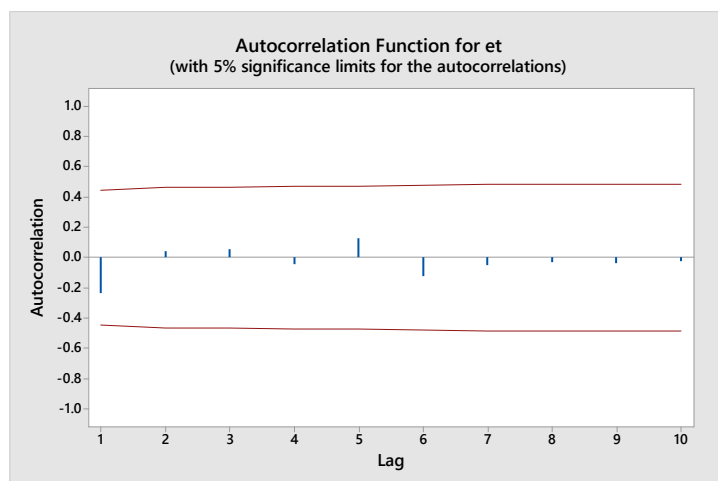


Figure 4.2: Autocorrelation function for residuals

จาก Autocorrelation function ของ residuals สามารถสังเกตเห็นได้ว่า residuals มีการกระจายแบบ random ดังนั้นสามารถสรุปได้ว่า Simple linear regression model นั้น adequate

4.1.2 Simple exponential smoothing method

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

Holt-winters exponential smoothing without trend and without seasonal component.

```
call:
Holtwinters(x = srts, beta = FALSE, gamma = FALSE)
```

```
Smoothing parameters:
alpha: 0.9999201
beta : FALSE
gamma: FALSE
```

```
Coefficients:
      [,1]
a 7.369942
```

Figure 4.2: Smoothing parameters optimization from RStudio

จากการที่ได้ทำ optimization ใน RStudio เพื่อหาค่า smoothing parameters หรือก็คือค่า alpha เพื่อใช้ในสมการพยากรณ์ของวิธี Simple exponential smoothing method เพื่อให้ได้ผลลัพธ์ของการพยากรณ์ที่มีความแม่นยำมากที่สุด ซึ่งค่า alpha ที่ดีที่สุดที่ได้จาก RStudio มีค่าเท่ากับ 0.9999201 ดังนั้นเราจึงนำค่า alpha นี้ไปแทนใน test set เพื่อหาค่าพยากรณ์ เพื่อทำการเปรียบเทียบ forecast errors กับตัว train set สำหรับใช้ในการพิจารณา overfitting ของสมการพยากรณ์

year	suicide rate/100k	\hat{Y}_t	e_t	e_t^2	abs	abs/ Y_t	e_t/Y_t
2540	6.92						
2541	8.12	6.920000	1.20	1.4400	1.2000	0.1478	0.1478
2542	8.59	8.119925	0.47	0.2210	0.4701	0.0547	0.0547
2543	8.40	8.589971	-0.19	0.0361	0.1900	0.0226	-0.0226
2544	7.74	8.400012	-0.66	0.4356	0.6600	0.0853	-0.0853
2545	7.84	7.740041	0.10	0.0100	0.1000	0.0127	0.0127
2546	7.13	7.839994	-0.71	0.5041	0.7100	0.0996	-0.0996
2547	6.87	7.130044	-0.26	0.0676	0.2600	0.0379	-0.0379
2548	6.34	6.870016	-0.53	0.2809	0.5300	0.0836	-0.0836
2549	5.77	6.340033	-0.57	0.3249	0.5700	0.0988	-0.0988
2550	5.97	5.770035	0.20	0.0400	0.2000	0.0335	0.0335
2551	5.98	5.969988	0.01	0.0001	0.0100	0.0017	0.0017
2552	5.97	5.979999	-0.01	0.0001	0.0100	0.0017	-0.0017
2553	5.90	5.970001	-0.07	0.0049	0.0700	0.0119	-0.0119
2554	6.03	5.900004	0.13	0.0169	0.1300	0.0216	0.0216
2555	6.20	6.029992	0.17	0.0289	0.1700	0.0274	0.0274
2556	6.08	6.199989	-0.12	0.0144	0.1200	0.0197	-0.0197
2557	6.07	6.080007	-0.01	0.0001	0.0100	0.0016	-0.0016
2558	6.47	6.070001	0.40	0.1600	0.4000	0.0618	0.0618
2559	6.35	6.469975	-0.12	0.0144	0.1200	0.0189	-0.0189

Table 4.4: Train set of Simple exponential smoothing method

year	suicide rate/100k	\hat{Y}_t	e_t	e_t^2	abs	abs/ Y_t	e_t/Y_t
2560	6.03	6.350010	-0.32	0.1024	0.3200	0.0531	-0.0531
2561	6.32	6.030026	0.29	0.0841	0.2900	0.0459	0.0459
2562	6.64	6.319977	0.32	0.1024	0.3200	0.0482	0.0482
2563	7.37	6.639974	0.73	0.5329	0.7300	0.0991	0.0991

Table 4.5: Test set of Simple exponential smoothing method

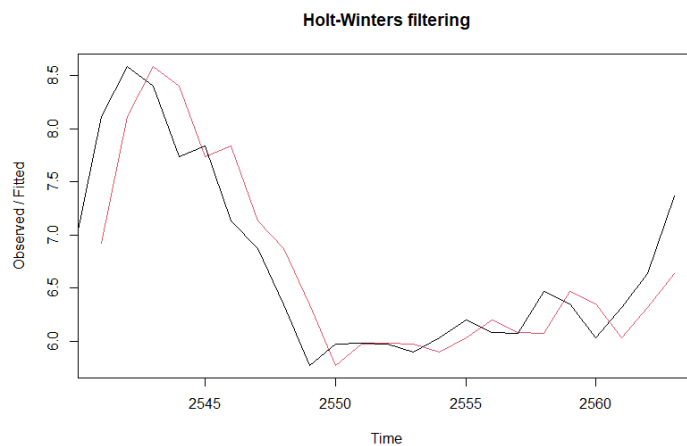


Figure 4.3: Train plot for Simple exponential smoothing method

Forecast error	Train set	Test set
MAD	0.3121	0.4150
MSE	0.1895	0.2055
RMSE	0.4353	0.4533
MAPE	4.4356%	6.1550%
MPE	-0.6332%	3.5016%

Table 4.6: Forecast errors of train versus test set

จากผลของการเปรียบเทียบ forecast error ของตัว train set กับ test set จะเห็นได้ว่าค่าแนวโน้มของค่า forecast error แต่ละตัว ของตัว test set มีค่ามากกว่าของตัว train set เพียงเล็กน้อย โดยทางผู้จัดทำได้พิจารณาจากค่า MAPE เป็นหลัก ดังนั้นจึงสามารถสรุปได้ว่าผลของ overfitting นั้นไม่ significant

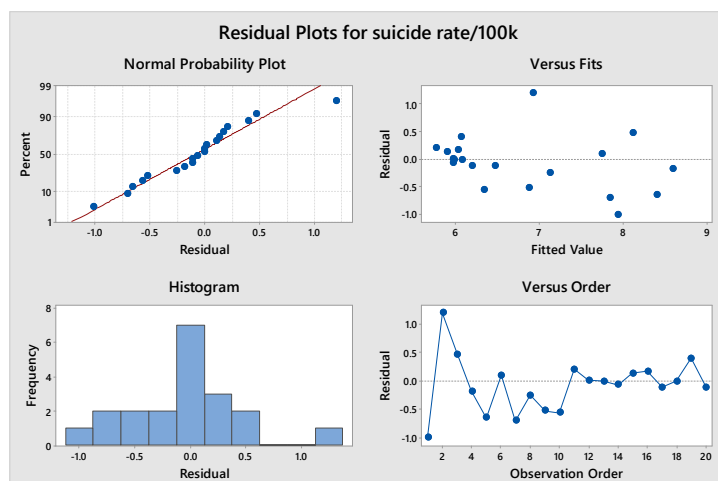


Figure 4.4: Residual plots of Simple exponential smoothing method

จาก Residual plots ที่ได้จาก Minitab เราสามารถสรุปได้ว่า residuals ที่ได้จากการถดถอยของ Simple linear regression model นั้นมีการกระจายตัวแบบปกติ มีค่าเฉลี่ยคงที่ มีความแปรปรวนร่วมกัน และเป็นอิสระต่อกัน

4.1.3 Holt's linear method

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} - T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$\hat{Y}_{t+p} = L_t + pT_t$$

Holt-winters exponential smoothing with trend and without seasonal component.

Call:
Holtwinters(x = srts, gamma = FALSE)

Smoothing parameters:
alpha: 0.9402521
beta : 0.8664106
gamma: FALSE

Coefficients:
[,1]
a 7.3447663
b 0.6628002

Figure 4.5: Smoothing parameters optimization from RStudio

จากการที่เราได้ทำ optimization ใน RStudio เพื่อหาค่า smoothing parameters ซึ่งได้แก่ค่า alpha และ beta เพื่อหาค่าที่ดีที่สุด ในการทำให้ค่าพยากรณ์ที่ได้มีความแม่นยำมากที่สุด ผลที่ได้คือ ค่า alpha มีค่าเท่ากับ 0.9402521 และ beta มีค่าเท่ากับ 0.8664106 หลังจากนั้น จึงได้นำค่าไปแทนเพื่อหาค่าพยากรณ์ใน test set เพื่อหา forecast errors โดยนำไปเปรียบเทียบกับตัว train set เพื่อพิจารณา overfitting

year	suicide rate/100k	L_t	T_t	\hat{Y}_{t+1}	e_t	e_t^2	abs	abs ²	abs/ Y_t	e_t/Y_t
2540	6.92	6.92	0	6.92	0	0	0	0	0	0
2541	8.12	8.048303	0.977573	6.92	1.2	1.44	1.2	1.44	0.147783251	0.147783
2542	8.59	8.616043	0.62249	9.025876	-0.43588	0.189988	0.435876	0.1899877	0.050742233	-0.05074
2543	8.4	8.450101	-0.06062	9.238532	-0.83853	0.703136	0.838532	0.70313619	0.099825258	-0.09983
2544	7.74	7.778805	-0.58971	8.389485	-0.64948	0.42183	0.649485	0.42183014	0.083912728	-0.08391
2545	7.84	7.80111	-0.05946	7.18909	0.65091	0.423683	0.65091	0.42368331	0.083024183	0.083024
2546	7.13	7.166545	-0.55774	7.741654	-0.61165	0.374121	0.611654	0.37412115	0.085786036	-0.08579
2547	6.87	6.854394	-0.34496	6.608809	0.261191	0.068221	0.261191	0.06822067	0.038019049	0.038019
2548	6.34	6.350123	-0.48299	6.509436	-0.16944	0.028709	0.169436	0.0287086	0.026724942	-0.02672
2549	5.77	5.775804	-0.56212	5.867135	-0.09713	0.009435	0.097135	0.00943521	0.016834488	-0.01683
2550	5.97	5.924812	0.054009	5.213685	0.756315	0.572013	0.756315	0.57201287	0.126685984	0.126686
2551	5.98	5.97993	0.05497	5.978821	0.001179	1.39E-06	0.001179	1.3904E-06	0.000197183	0.000197
2552	5.97	5.973878	0.0021	6.034899	-0.0649	0.004212	0.064899	0.00421191	0.010870894	-0.01087
2553	5.9	5.904539	-0.05979	5.975977	-0.07598	0.005773	0.075977	0.00577258	0.01287754	-0.01288
2554	6.03	6.018931	0.091122	5.844745	0.185255	0.03432	0.185255	0.03431951	0.030722263	0.030722
2555	6.2	6.194626	0.164397	6.110054	0.089946	0.00809	0.089946	0.00809032	0.014507453	0.014507
2556	6.08	6.096671	-0.06291	6.359022	-0.27902	0.077854	0.279022	0.07785354	0.045891855	-0.04589
2557	6.07	6.067835	-0.03339	6.033763	0.036237	0.001313	0.036237	0.00131308	0.005969771	0.00597
2558	6.47	6.443977	0.321433	6.034447	0.435553	0.189706	0.435553	0.18970616	0.06731881	0.067319
2559	6.35	6.37482	-0.01698	6.76541	-0.41541	0.172565	0.41541	0.1725651	0.065418828	-0.06542
2560	6.03	6.049588	-0.28405	6.357842	-0.32784	0.10748	0.327842	0.10748018	0.054368441	-0.05437
2561	6.32	6.286872	0.167639	5.765535	0.554465	0.307431	0.554465	0.3074311	0.087731756	0.087732
2562	6.64	6.628917	0.318747	6.454511	0.185489	0.034406	0.185489	0.03440621	0.027935105	0.027935
2563	7.37	7.344766	0.6628	6.947664	0.422336	0.178368	0.422336	0.17836764	0.057304739	0.057305

Table 4.7: Train set of Holt's linear method

year	suicide rate/100k	L_t	T_t	\hat{Y}_{t+1}	e_t	e_t^2	abs	abs ²	abs/ Y_t	e_t/Y_t
2560	6.03	6.049588	-0.28405	6.357842	-0.32784	0.10748	0.327842	0.10748029	0.05436847	-0.05437
2561	6.32	6.286872	0.167639	5.765536	0.554464	0.307431	0.554464	0.30743083	0.087731717	0.087732
2562	6.64	6.628917	0.318746	6.454511	0.185489	0.034406	0.185489	0.0344063	0.027935145	0.027935
2563	7.37	7.344766	0.6628	6.947664	0.422336	0.178368	0.422336	0.17836779	0.057304764	0.057305

Table 4.8: Test set of Holt's linear method

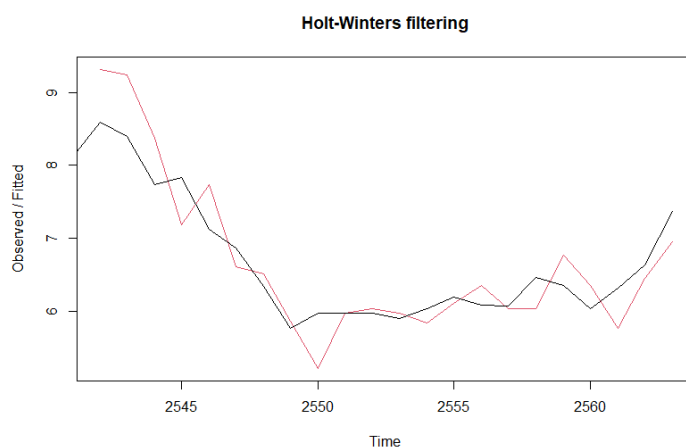


Figure 4.6: Train plot for Holt's linear method

Forecast error	Train set	Test set
MAD	0.3627	0.3725
MSE	0.2362	0.1569
RMSE	0.4861	0.3961
MAPE	5.0656%	5.6835%
MPE	0.0767%	2.9651%

Table 4.9: Forecast errors of train versus test set

จากผลของการเปรียบเทียบ forecast error ของตัว train set กับ test set จะเห็นได้ว่าค่าแนวโน้มของค่า forecast error แต่ละตัว ของตัว test set มีค่ามากกว่าของตัว train set เพียงเล็กน้อย โดยทางผู้จัดทำได้พิจารณาจากค่า MAPE เป็นหลัก ดังนั้นจึงสามารถสรุปได้ว่าผลของ overfitting นั้นไม่ significant

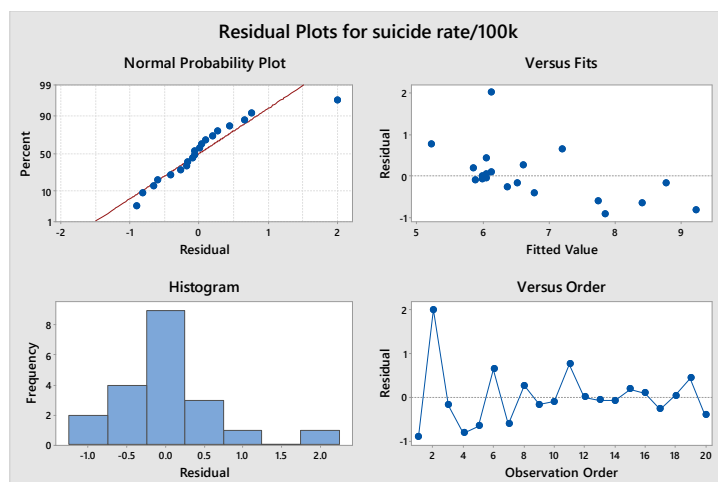


Figure 4.7: Residual plots of Holt's linear method

จาก Residual plots ที่ได้จาก Minitab เราสามารถสรุปได้ว่า residuals ที่ได้จากการถดถอยของ Simple linear regression model นั้นมีการกระจายตัวแบบปกติ มีค่าเฉลี่ยคงที่ มีความแปรปรวนร่วมกัน และเป็นอิสระต่อกัน

Forecast error	Naïve (trend)	Simple Expo	Holt's
MAD	0.3778	0.3121	0.3627
MSE	0.2172	0.1895	0.2362
RMSE	0.4660	0.4353	0.4861
MAPE	5.4156%	4.4356%	5.0656%
MPE	-0.8319%	-0.6332%	0.0767%

Table 4.10: Measurement error of time series method

จากการวิเคราะห์ค่า forecast errors พบว่าวิธีของ Simple exponential smoothing method ให้ค่า forecast errors ที่ต่ำกว่าอีก 2 วิธี อย่างเห็นได้ชัด ดังนั้นทางผู้จัดทำจึงได้เลือกวิธี Simple exponential smoothing method ดังกล่าว ในการใช้เป็นแบบจำลองพยากรณ์ของวิธี Time-series method

4.2 Regression with time-series method

Correlations

	suicide rate/100k
growth rate	0.848
life expectancy	-0.721
population	-0.589
inflation rate	-0.058
manufacturing	-0.791
GDP per capita	-0.754
GNI per capita	-0.726
no.of tourists	-0.553
trade balance	-0.128
imports	-0.793
exports	-0.787
poverty rate	0.823
hungers	0.873
unemployment rate	0.905
youth rate	0.780
participation rate	0.620
crime rate/100k	0.647
health per capita	-0.713
military	-0.683

Figure 4.8: Correlations of predictors to suicide rate/100k

4.2.1 Simple linear regression

จาก Correlations ที่ได้ ผู้จัดทำจึงได้เลือกตัวแปรที่มีค่า correlation สูงที่สุด ที่เทียบกับ อัตราการฆ่าตัวตาย ที่เป็นตัวแปรตามที่เราสนใจ ดังนั้นเราจึงได้เลือกตัวแปร Unemployment rate ที่มีค่า correlation เท่ากับ 0.905 นำมาเป็นตัวแปรที่ใช้ในการทำนายในการสร้างสมการ Simple linear regression

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

4.2.1.1 Regression analysis

TRAIN SET

Regression Analysis: suicide rate/100k versus unemployment rate

Backward Elimination of Terms

α to remove = 0.1

Regression Equation

suicide rate/100k = 5.428 + 0.6908 unemployment rate

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.428	0.171	31.77	0.000	
unemployment rate	0.6908	0.0767	9.00	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.401498	81.83%	80.82%	78.02%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	13.0672	13.0672	81.06	0.000
unemployment rate	1	13.0672	13.0672	81.06	0.000
Error	18	2.9016	0.1612		
Lack-of-Fit	16	2.8816	0.1801	17.96	0.054
Pure Error	2	0.0201	0.0100		
Total	19	15.9688			

Durbin-Watson Statistic

Durbin-Watson Statistic = 1.19957

Figure 4.9: Simple linear regression analysis

จาก Regression analysis ที่ได้จาก Minitab พบว่าได้ค่า $\beta_0 = 5.428$ และ $\beta_1 = 0.6908$ ซึ่งค่า p-value ของ β_1 นั้นมีค่าน้อยกว่า α ที่มีค่าเท่ากับ 0.05 หมายความว่าสัมประสิทธิ์นั้นมีค่าแตกต่างจาก 0 อย่างมีนัยสำคัญ และผลปรากฏว่าได้ค่าความแม่นยำของสมการถดถอย โดยพิจารณาจากค่า R-sq(adj) อยู่ที่ 80.82% ซึ่งอยู่ในเกณฑ์ที่มีความแม่นยำสูง

$$Y = 5.428 + 0.6908X$$

4.2.1.2 Durbin-Watson test

DW	n	k	d_L	d_U
1.19957	20	1	1.20	1.41

Table 4.11: Durbin-Watson test of Simple linear regression model

และจาก Regression analysis ที่ได้ พบว่าได้ค่า Durbin-Watson Statistic มีค่าเท่ากับ 1.19957 ซึ่งมีค่าน้อยกว่าค่า d_L ที่ได้จากตาราง ดังนั้นจึง accept null hypothesis และสามารถสรุปได้ว่ามี first-order serial correlation ปรากฏใน Simple linear regression model

4.2.1.3 Residual plots and ACF for residuals

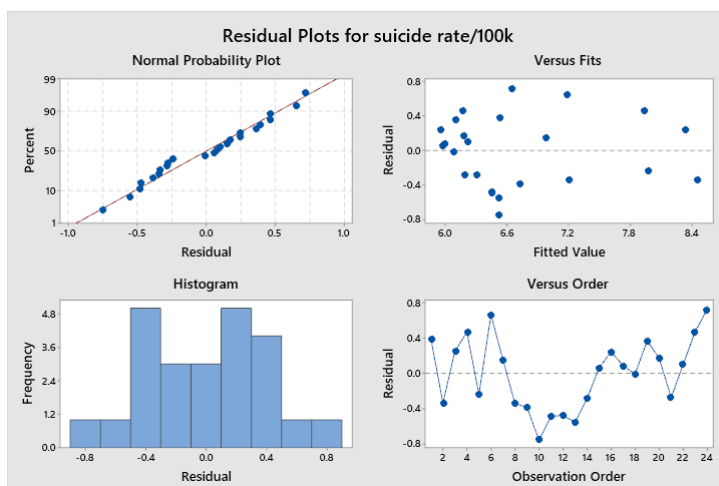


Figure 4.10: Residual plots of Simple linear regression model

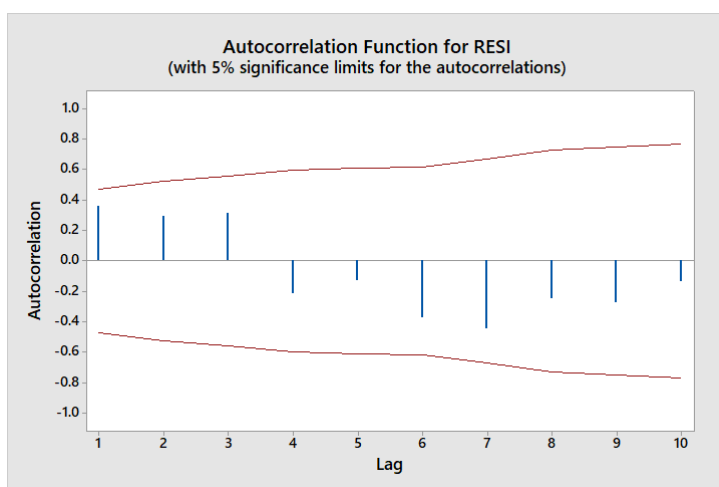


Figure 4.11: Autocorrelation function for residuals

จาก Residual plots ที่ได้จาก Minitab เราสามารถสรุปได้ว่า residuals ที่ได้จากสมการถดถอยของ Simple linear regression model นั้นมีการกระจายตัวแบบปกติ มีค่าเฉลี่ยคงที่ มีความแปรปรวนร่วมกัน และเป็นอิสระต่อกัน และจาก Autocorrelation function ของ residuals สามารถสังเกตเห็นได้ว่า residuals มีการกระจายตัวแบบ random ดังนั้นจึงสามารถสรุปได้ว่า Simple linear regression model ที่ได้นั้น adequate

4.2.1.4 Overfitting check

Forecast error	Train set	Test set
MAD	0.3424	0.4222
MSE	0.1497	0.1040
RMSE	0.3809	0.4896
MAPE	5.0980%	6.6553%
MPE	0.0121%	4.9463%

Table 4.12: Forecast errors of train and test set

จากผลของการเปรียบเทียบ forecast error ของตัว train set กับ test set จะเห็นได้ว่าค่าแนวโน้มของค่า forecast error แต่ละตัว ของตัว test set มีค่ามากกว่าของตัว train set เพียงเล็กน้อย โดยทางผู้จัดทำได้พิจารณาจากค่า MAPE เป็นหลัก ดังนั้นจึงสามารถสรุปได้ว่าผลของ overfitting นั้นไม่ significant

4.2.2 Multiple linear regression

จาก Correlation matrix ที่ทางผู้จัดทำได้จาก Minitab ในขั้นตอนก่อนหน้า ตามที่ผู้จัดทำได้ตั้งเกณฑ์สำหรับการเพิ่มตัวแปรเพิ่มเข้าไปในสมการถดถอยเพื่อใช้ในการพิจารณา โดยถ้าค่า correlation ของตัวแปรนั้น ๆ มีค่ามากกว่า 0.60 ก็จะนำตัวแปรนั้นมาพิจารณา ตาม remaining predictors ใน Table: 3.9 โดยมีจำนวนทั้งหมด 15 ตัวแปร เพื่อนำไปใช้ใน Minitab โดยใช้วิธี Stepwise regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

4.2.2.1 Regression analysis

Stepwise Selection of Terms

α to enter = 0.1, α to remove = 0.1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	15.1117	3.77793	66.12	0.000
growth rate	1	0.4581	0.45811	8.02	0.013
hungers	1	0.7844	0.78440	13.73	0.002
unemployment rate	1	0.3127	0.31267	5.47	0.034
participation rate	1	1.5445	1.54449	27.03	0.000
Error	15	0.8571	0.05714		
Total	19	15.9688			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.239039	94.63%	93.20%	88.22%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	9.67	1.01	9.62	0.000	
growth rate	1.849	0.653	2.83	0.013	11.07
hungers	0.2155	0.0582	3.71	0.002	12.19
unemployment rate	0.240	0.103	2.34	0.034	5.04
participation rate	-0.1485	0.0286	-5.20	0.000	6.49

Regression Equation

suicide rate/100k = 9.67 + 1.849 growth rate + 0.2155 hungers + 0.240 unemployment rate
- 0.1485 participation rate

Durbin-Watson Statistic

Durbin-Watson Statistic = 2.64957

Figure 4.12: Multiple regression analysis

จาก Regression analysis ผลปรากฏว่าได้ค่า $\beta_0 = 9.67$, $\beta_1 = 1.849$, $\beta_2 = 0.2155$, $\beta_3 = 0.240$ และ $\beta_4 = -0.1485$ ซึ่งค่า p-value ของ β_1 , β_2 , β_3 และ β_4 นั้นมีค่าน้อยกว่า α ซึ่งมีค่าเท่ากับ 0.05 ซึ่งหมายความว่าค่าสัมประสิทธิ์ที่ได้นั้นมีค่าต่างจาก 0 อย่างมีนัยสำคัญ และพบว่าได้ค่าความแม่นยำ โดยพิจารณาจากค่า R-sq(adj) โดยมีค่าเท่ากับ 93.20% ซึ่งอยู่ในเกณฑ์ที่มีความแม่นยำสูง

$$Y = 9.67 + 1.849X_1 + 0.2155X_2 + 0.240X_3 - 0.1485X_4$$

4.2.2.2 Durbin-Watson test

DW	n	k	d_L	d_U
2.64957	20	3	1.00	1.68

Table 4.13: Durbin-Watson test of Multiple linear regression model

และจาก Regression analysis ที่ได้ พบว่าได้ค่า Durbin-Watson Statistic มีค่าเท่ากับ 2.64957 ซึ่งมีค่าน้อยกว่าค่า d_U ที่ได้จากตารางที่ 5% significance level ดังนั้นจึง reject null hypothesis และสามารถสรุปได้ว่าไม่ปรากฏ first-order serial correlation ที่ significant ภายในสมการถดถอย Multiple linear regression model

4.2.2.3 Residual plots and ACF for residuals

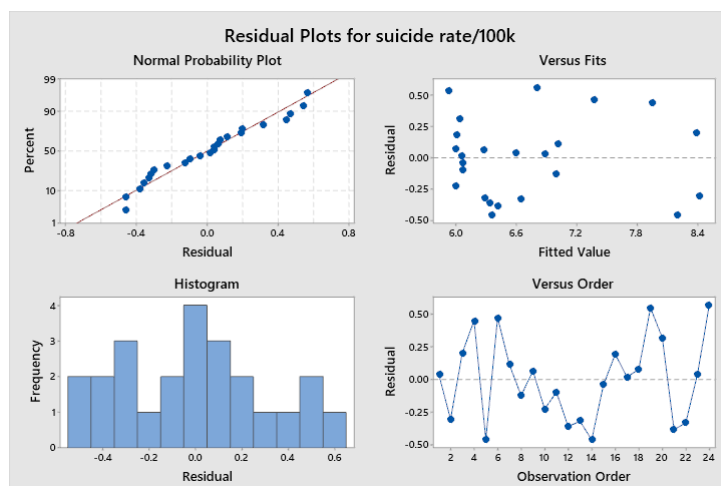


Figure 4.13: Residuals for Multiple linear regression model

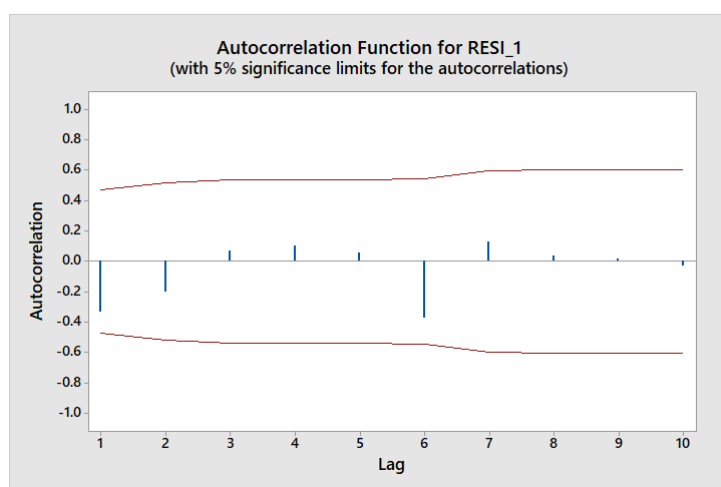


Figure 4.14: Autocorrelation function for residuals

จาก Residual plots ที่ได้จาก Minitab เราสามารถสรุปได้ว่า residuals ที่ได้จากสมการถดถอยของ Multiple linear regression model นั้นมีการกระจายตัวแบบปกติ มีค่าเฉลี่ยคงที่ มีความแปรปรวนร่วมกัน และเป็นอิสระต่อกัน และจาก Autocorrelation function ของ residuals สามารถสังเกตเห็นได้ว่า residuals มีการกระจายตัวแบบ random ดังนั้นจึงสามารถสรุปได้ว่า Multiple linear regression model ที่ได้นั้น adequate

4.2.2.4 Overfitting check

Forecast error	Train set	Test set
MAD	0.1698	0.4312
MSE	0.0451	0.0749
RMSE	0.2070	0.5411
MAPE	2.5788%	6.6952%
MPE	0.0096%	2.7108%

Table 4.14: Forecast errors of train versus test set

จากผลของการเปรียบเทียบ forecast error ของตัว train set กับ test set จะเห็นได้ว่าค่าแนวโน้มของค่า forecast error แต่ละตัว ของตัว test set มีค่ามากกว่าของตัว train set เพียงเล็กน้อย โดยทางผู้จัดทำได้พิจารณาจากค่า MAPE เป็นหลัก ดังนั้นจึงสามารถสรุปได้ว่าผลของ overfitting นั้นไม่ significant

4.3 The BOX-JENKINS (ARIMA)

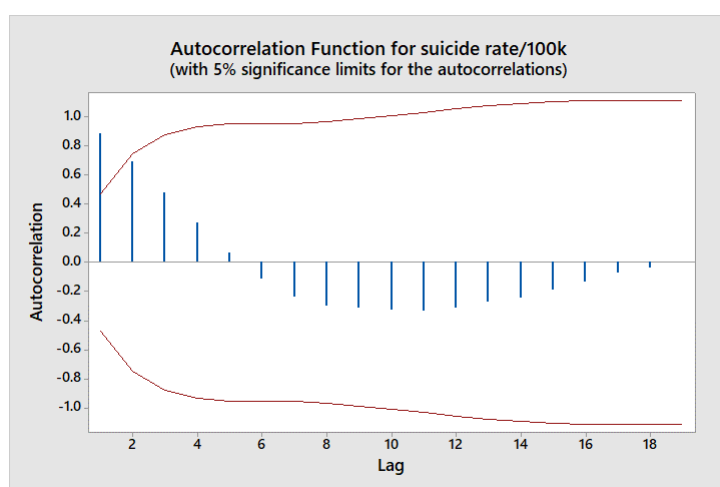


Figure 4.15: Autocorrelation function for suicide rate/100k

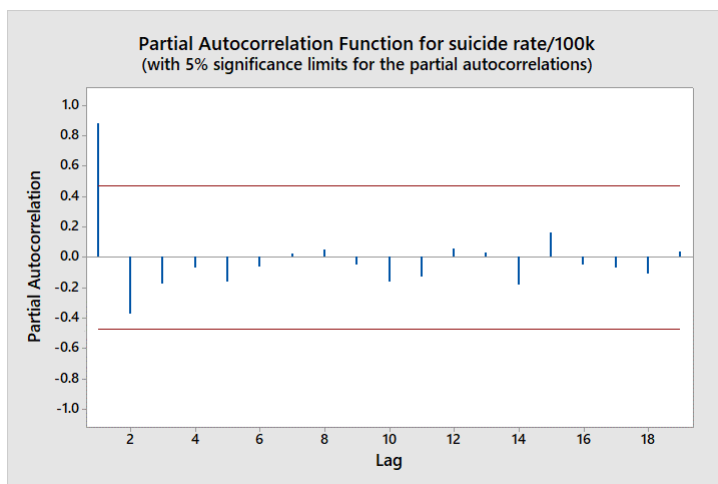


Figure 4.16: Partial autocorrelation function for suicide rate/100k

จากผลของ Autocorrelation และ Partial autocorrelation function ที่ได้จาก Minitab เราจะเห็นว่าจากกราฟของ ACF ค่อนข้างที่จะมีลักษณะเป็น trend และค่อย ๆ มีแนวโน้มลดลง ดังนั้นทางผู้จัดทำ จึงได้ทำการพิจารณาข้อมูลอัตราการฆ่าตัวตายเป็น non-stationary data และได้มีการพิจารณาในการทำ differencing ประกอบ และจากกราฟของ PACF ที่ได้ จะเห็นได้ว่าที่ time lag 1 นั้น significant แล้วก็ได้ cut off หายไป ดังนั้นทางผู้จัดทำจึงตัดสินใจที่จะพิจารณา ARIMA models ทั้งหมด 3 รูปแบบ ได้แก่ ARIMA(0,1,0), ARIMA(1,0,0) และ ARIMA(2,0,0)

4.3.1 ARIMA(0,1,0)

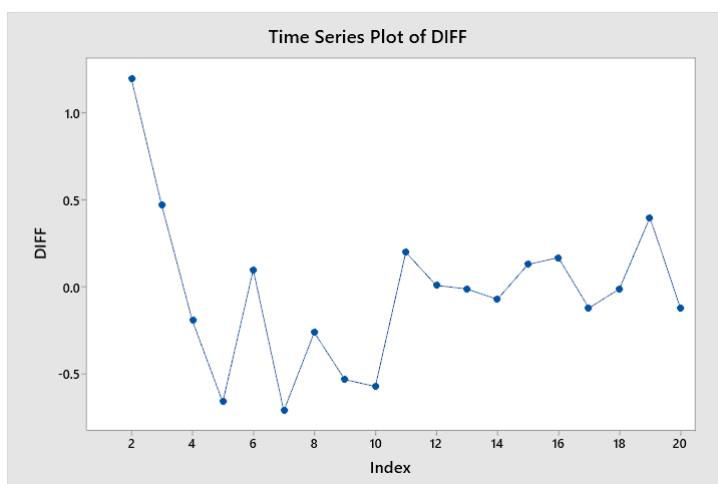


Figure 4.17: Time-series plot of differenced suicide rate series

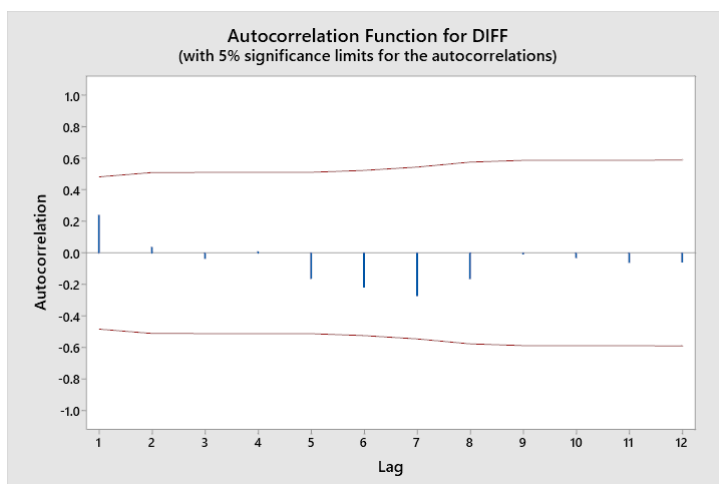


Figure 4.18: Autocorrelation function for differenced series

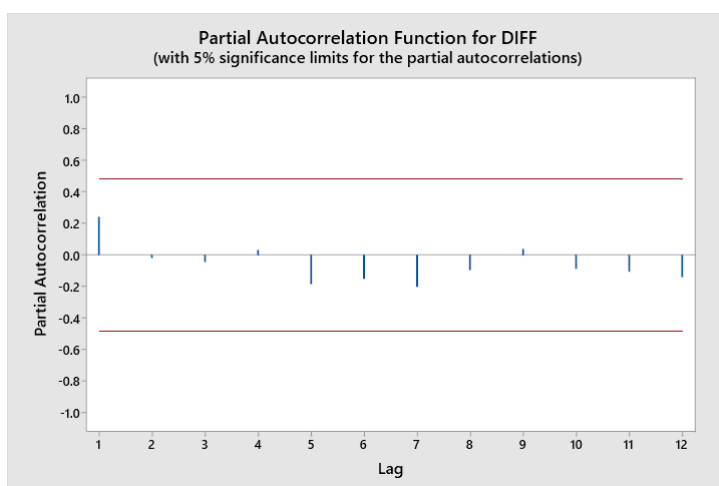


Figure 4.19: Partial autocorrelation function for differenced series

จากกราฟของ ACF ของอัตราการฆ่าตัวตายรายปี ที่มีแนวโน้มค่อนข้างเป็น trend และค่อย ๆ ลดลง ดังนั้นทางผู้จัดทำจึงได้ลองทำการ differencing 1 ครั้ง ($d=1$) เพื่อให้ pattern ของข้อมูลมีลักษณะเข้าใกล้ stationary data และจากการพิจารณารูป ACF และ PACF ของข้อมูลหลังจากที่ผ่านการทำ differencing จะเห็นได้ว่าไม่ปรากฏความสัมพันธ์ใด ๆ ที่เกี่ยวข้องกับ Autoregressive (AR) และ Moving average (MA) ดังนั้นจึงได้ทำการพิจารณา ARIMA(0,1,0) หรือที่เรียกว่า Random walk

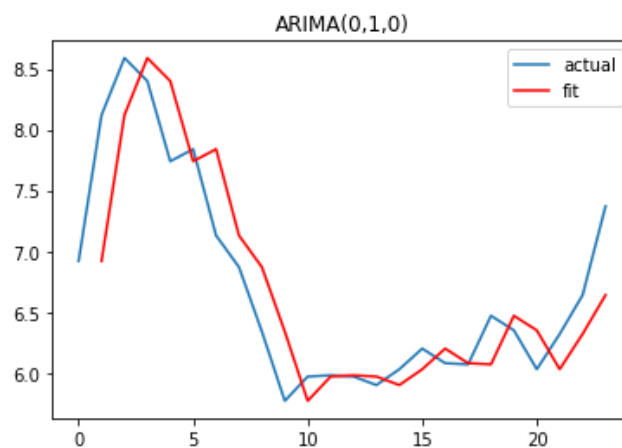


Figure 4.20: ARIMA(0,1,0) actual versus fitted

```

=====
SARIMAX Results
=====
Dep. Variable:    suicide rate/100k    No. Observations:    24
Model:            ARIMA(0, 1, 0)       Log Likelihood       -13.672
Date:             Sat, 20 Nov 2021     AIC                  29.345
Time:             07:11:14             BIC                  30.480
Sample:           0                    HQIC                 29.630
Covariance Type:  opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
sigma2         0.1922     0.049       3.899     0.000     0.096     0.289
=====
Ljung-Box (L1) (Q):           1.99    Jarque-Bera (JB):           1.44
Prob(Q):                     0.16    Prob(JB):                 0.49
Heteroskedasticity (H):       0.34    Skew:                     0.55
Prob(H) (two-sided):          0.15    Kurtosis:                 3.55
=====

```

Figure 4.21: ARIMA(0,1,0) SARIMAX results

AIC	29.345	Forecast error	Train set	Test set
BIC	30.480	MAD	0.3121	0.4150
HQIC	29.630	MSE	0.1992	0.0504
		RMSE	0.4353	0.5411
		MAPE	4.4441%	6.5354%
		MPE	-0.2686%	3.7795%

Table 4.15: Criterion and Forecast errors of train versus test set

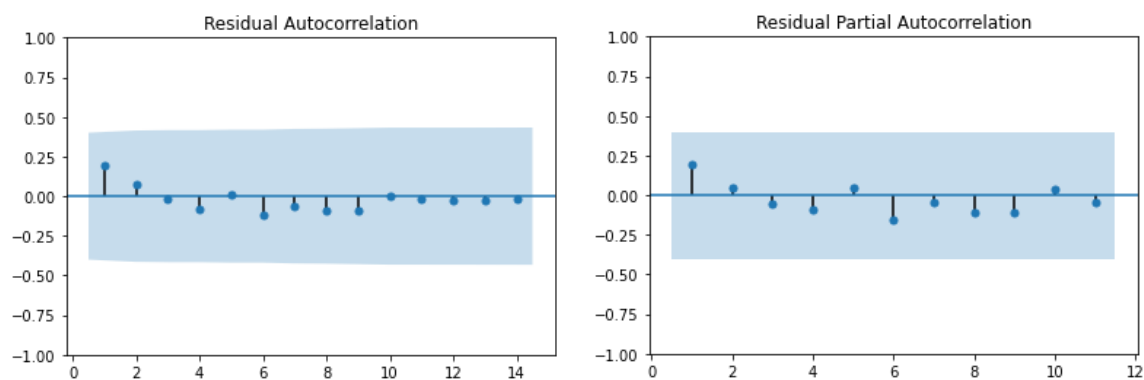


Figure 4.22: ACF and PACF for residuals

จากการที่ได้พิจารณา และวิเคราะห์ overfitting จากการเปรียบเทียบระหว่าง forecast errors ของตัว train set กับ test set ผลปรากฏว่าค่า forecast errors มีค่าแตกต่างกันเพียงเล็กน้อย ดังนั้นจึงสามารถสรุปได้ว่าไม่ปรากฏ overfitting ภายใน ARIMA(0,1,0) model และจาก ACF และ PACF ของ residuals ที่ได้จาก Statmodels จะสังเกตเห็นได้ว่า residuals มีการกระจายแบบ random ไม่หลงเหลือ pattern ใด ๆ ในของชุดข้อมูล ดังนั้น ARIMA(0,1,0) model นี้จึง adequate

4.3.2 ARIMA(1,0,0)

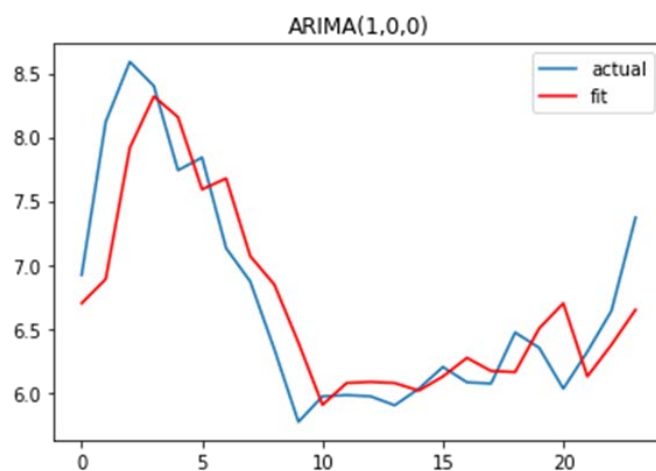


Figure 4.23: ARIMA(1,0,0) actual versus fitted

```

=====
SARIMAX Results
=====
Dep. Variable:    suicide rate/100k    No. Observations:    24
Model:            ARIMA(1, 0, 0)        Log Likelihood        -13.743
Date:             Sat, 20 Nov 2021      AIC                   33.486
Time:             07:16:41              BIC                   37.020
Sample:           0                     HQIC                  34.423
Covariance Type:  opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
const         6.8504     0.680     10.069     0.000     5.517     8.184
ar.L1         0.8488     0.122      6.967     0.000     0.610     1.088
sigma2        0.1745     0.048      3.624     0.000     0.080     0.269
=====
Ljung-Box (L1) (Q):                2.90    Jarque-Bera (JB):                5.25
Prob(Q):                           0.09    Prob(JB):                       0.07
Heteroskedasticity (H):             0.36    Skew:                           0.99
Prob(H) (two-sided):                0.17    Kurtosis:                       4.15
=====

```

Figure 4.24: ARIMA(1,0,0) SARIMAX results

$$\hat{Y}_t = 0.9599 + 1.3287Y_{t-1} - 0.5264Y_{t-2}$$

AIC	33.486	Forecast error	Train set	Test set
BIC	37.020	MAD	0.3015	0.3755
HQIC	34.423	MSE	0.1813	0.0499
		RMSE	0.4163	0.4787
		MAPE	4.3440%	5.8310%
		MPE	-0.1896%	2.0087%

Table 4.16: Criterion and Forecast errors of train versus test set

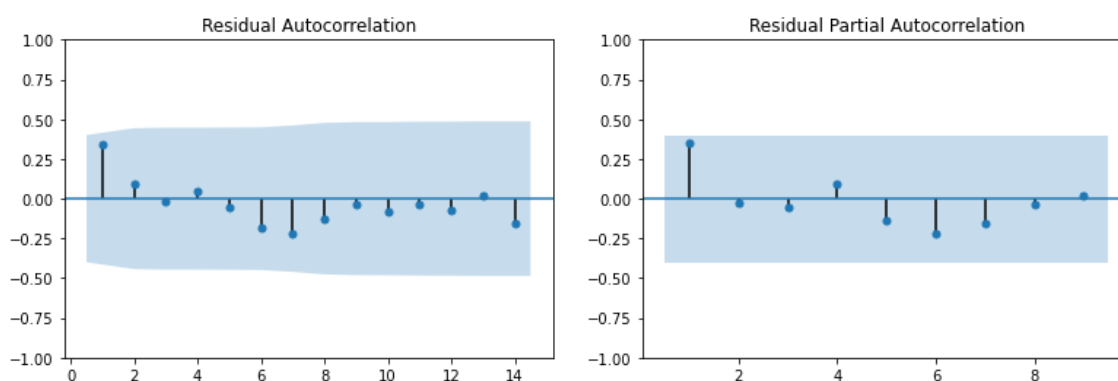


Figure 4.25: ACF and PACF for residuals

จากการที่ได้พิจารณา และวิเคราะห์ overfitting จากการเปรียบเทียบระหว่าง forecast errors ของตัว train set กับ test set ผลปรากฏว่าค่า forecast errors มีค่าแตกต่างกันเพียงเล็กน้อย ดังนั้นจึงสามารถสรุปได้ว่าไม่ปรากฏ overfitting ภายใน ARIMA(1,0,0) model และจาก ACF และ PACF ของ residuals ที่ได้จาก Statmodels จะสังเกตเห็นได้ว่า residuals มีการกระจายแบบ random ไม่หลงเหลือ pattern ใด ๆ ในของชุดข้อมูล ดังนั้น ARIMA(1,0,0) model นี้จึง adequate

4.3.2 ARIMA(2,0,0)

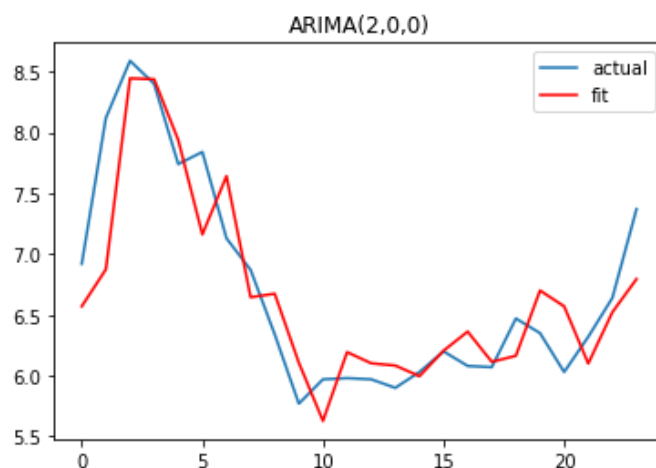


Figure 4.26: ARIMA(2,0,0) actual versus fitted

SARIMAX Results						
=====						
Dep. Variable:	suicide rate/100k	No. Observations:	24			
Model:	ARIMA(2, 0, 0)	Log Likelihood	-11.079			
Date:	Sat, 20 Nov 2021	AIC	30.157			
Time:	07:14:38	BIC	34.869			
Sample:	0	HQIC	31.407			
	- 24					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	6.6976	0.463	14.454	0.000	5.789	7.606
ar.L1	1.3580	0.229	5.918	0.000	0.908	1.808
ar.L2	-0.5714	0.168	-3.409	0.001	-0.900	-0.243
sigma2	0.1347	0.048	2.785	0.005	0.040	0.229
=====						
Ljung-Box (L1) (Q):		0.69	Jarque-Bera (JB):		3.47	
Prob(Q):		0.41	Prob(JB):		0.18	
Heteroskedasticity (H):		0.45	Skew:		0.91	
Prob(H) (two-sided):		0.28	Kurtosis:		3.40	
=====						

Figure 4.27: ARIMA(2,0,0) SARIMAX results

$$\hat{Y}_t = 1.2989 + 1.3287Y_{t-1} - 0.5264Y_{t-2}$$

AIC	30.157	Forecast error	Train set	Test set
BIC	34.869	MAD	0.2970	0.3600
HQIC	31.407	MSE	0.1646	0.0648
		RMSE	0.4031	0.4772
		MAPE	4.4331%	5.5874%
		MPE	0.5225%	2.8252%

Table 4.17: Criterion and Forecast errors of train versus test set

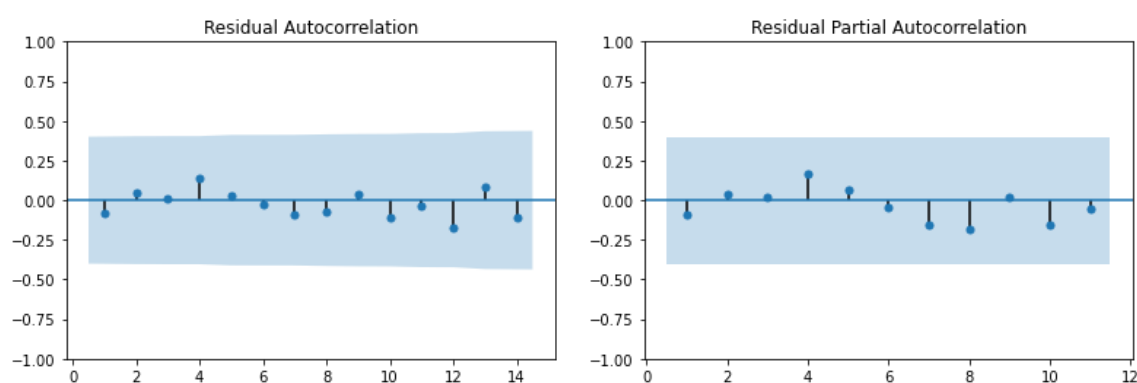


Figure 4.26: ACF and PACF for residuals

จากการที่ได้พิจารณา และวิเคราะห์ overfitting จากการเปรียบเทียบระหว่าง forecast errors ของตัว train set กับ test set ผลปรากฏว่าค่า forecast errors มีค่าแตกต่างกันเพียงเล็กน้อย ดังนั้นจึงสามารถสรุปได้ว่าไม่ปรากฏ overfitting ภายใน ARIMA(2,0,0) model และจาก ACF และ PACF ของ residuals ที่ได้จาก Statmodels จะสังเกตเห็นได้ว่า residuals มีการกระจายแบบ random ไม่หลงเหลือ pattern ใด ๆ ในของชุดข้อมูล ดังนั้น ARIMA(2,0,0) model นี้จึง adequate

Forecast error	ARIMA(0,1,0)	ARIMA(1,0,0)	ARIMA(2,0,0)
MAD	0.3121	0.3015	0.2970
MSE	0.1992	0.1813	0.1646
RMSE	0.4353	0.4163	0.4031
MAPE	4.441%	4.3440%	4.4331%
MPE	-0.2686%	-0.1896%	0.5225%

Table 4.18: Forecast errors of ARIMA models

Criterion	ARIMA(0,1,0)	ARIMA(1,0,0)	ARIMA(2,0,0)
AIC	29.345	33.486	30.157
BIC	30.480	37.020	34.869

Table 4.19: AIC and BIC of ARIMA models

เมื่อเปรียบเทียบ forecast errors ของ ARIMA models จะเห็นว่า ARIMA(2,0,0) นั้นค่อนข้างจะมี forecast errors ที่ต่ำเมื่อเทียบกับ ARIMA model อื่น ๆ ซึ่งสอดคล้องกับผลที่ได้จากการทำ Auto ARIMA ใน Python แต่ถ้าเปรียบเทียบในแง่ของ criterion ที่ใช้ ซึ่งก็คือ AIC และ BIC จะเห็นว่า ARIMA(0,1,0) นั้นมีค่า AIC และ BIC ที่ต่ำที่สุด เมื่อเทียบกับ ARIMA models ที่เหลือ ดังนั้นจึงสามารถเลือกใช้ ARIMA model ได้ก็ ได้ในการพยากรณ์อัตราการฆ่าตัวตายรายปี ขึ้นอยู่กับเงื่อนไข และสถานการณ์ที่ต้องใช้

4.4 Judgmental Forecasting

ผู้จัดทำมีความเห็นว่าในปี 2564 ประเทศไทยยังคงประสบปัญหามากมายที่เป็นผลมาจากการแพร่ระบาดของโรค COVID-19 โดยเฉพาะทางด้านเศรษฐกิจที่ถดถอยมาตั้งแต่ปี 2562 ที่เริ่มมีการระบาดเกิดขึ้น ซึ่งสอดคล้องกับการเพิ่มขึ้นของอัตราการฆ่าตัวตายที่เพิ่มมากขึ้นถึง 5% ระหว่างปี 2561 ถึง 2562 และอีก 11% ระหว่างปี 2562 ถึง 2563

ทางผู้จัดทำคาดว่าประเทศไทยยังจะได้รับผลกระทบจากการแพร่ระบาดของโรค COVID-19 ไปจนถึงปี 2565 จึงได้ทำการพยากรณ์อัตราการฆ่าตัวตายของปี 2564 ด้วยดุลยพินิจ จากการเฉลี่ยค่าอัตราการเพิ่มขึ้นของอัตราการฆ่าตัวตายตั้งแต่ปี 2560 จนถึงปี 2563 โดยนำปี 2560 มาพิจารณาด้วยเพื่อถ่วงไม่ให้เกิดการพยากรณ์ที่มีค่าที่สูงเกินไป โดยได้ค่าเฉลี่ยเป็น 7% ซึ่งจะได้ค่าพยากรณ์สำหรับอัตราการฆ่าตัวตายในปี 2564 เป็น 7.88 คน ต่อประชากรหนึ่งแสนคน

บทที่ 5

สรุปผลการดำเนินงาน และข้อเสนอแนะ

Forecast error	Simple Expo	Multiple Regression	ARIMA(0,1,0)	ARIMA(2,0,0)
MAD	0.3121	0.3358	0.3121	0.2970
MSE	0.1895	0.0997	0.1992	0.1646
RMSE	0.4353	0.3937	0.4353	0.4031
MAPE	4.4356%	5.0431%	4.441%	4.4331%
MPE	-0.6332%	-0.8884%	-0.2686%	0.5225%

Table 5.1: Comparison of forecast errors of each method

การพยากรณ์อัตราการฆ่าตัวตายสามารถทำได้หลายวิธีทั้ง Time-series method และ Causal method ซึ่งทางผู้จัดทำได้นำวิธีการต่าง ๆ มาใช้พยากรณ์ในชุดข้อมูลเดียวกัน แล้วนำไปหาค่า forecast errors เพื่อนำมาพิจารณาว่าวิธีการพยากรณ์แบบใดนั้นจะให้ผลการพยากรณ์ที่แม่นยำที่สุด ซึ่งเมื่อพิจารณาค่าของ MSE เพื่อที่จะไม่ให้เกิดค่าความคลาดเคลื่อนที่เกิดจากการพยากรณ์มากเกินไป (Penalizes large forecasting errors) พบว่าวิธีการพยากรณ์โดยการใช้สมการถดถอยแบบพหุคูณ หรือ Multiple regression model นั้นเป็นวิธีที่ให้ค่า MSE ต่ำที่สุด

สำหรับการต่อยอดงานวิจัยนี้ในอนาคต สำหรับ Time-series method สามารถหาวิธีการพยากรณ์ที่เหมาะสมกับรูปแบบของข้อมูลได้ซึ่งอาจเกิดจากการที่แบ่งส่วนของข้อมูลเป็นส่วน ๆ แล้ววิเคราะห์หารูปแบบข้อมูลในแต่ละส่วนเพื่อหาวิธีการพยากรณ์ที่เหมาะสมในแต่ละช่วงได้ และสำหรับวิธี Causal method สามารถเพิ่มข้อมูลเชิงคุณภาพ (Qualitative variable) เช่น ความคิดเห็นจากผู้เชี่ยวชาญ หรือสถานการณ์ COVID-19 ในขณะนั้นว่ามีความรุนแรงมากน้อยเพียงใด ซึ่งวิธีวิเคราะห์เกี่ยวกับข้อมูลเชิงคุณภาพก็จำเป็นต้องได้รับการศึกษา และวิจัยเพิ่มเติมต่อไป

รายการอ้างอิง

(n.d.). Retrieved November 24, 2021, from macrotrends:

<https://www.macrotrrends.net/countries/THA/thailand/>

(n.d.). Retrieved November 24, 2021, from National Statistical Office of Thailand:

<http://statbbi.nso.go.th/staticreport/page/sector/th/08.aspx>

(n.d.). Retrieved November 24, 2021, from Department of Mental Health, Thailand:

<https://www.dmh.go.th/report/suicide/>

John E. Hanke, Dean Wichern. (2014). *Business Forecasting Ninth Edition*. Pearson Education Limited.