



รายงานศึกษา

การพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบิน

ของสายการบิน United

โดยใช้เทคนิคการพยากรณ์และการประยุกต์

เสนอ

ผู้ช่วยศาสตราจารย์นพรัตน์ แป้นงาม

จัดทำโดย

นางสาวปวีณา โชติประเสริฐ รหัสนักศึกษา 65030292

นายชัชดานัย รุ้ทัน รหัสนักศึกษา 65030737

รายงานนี้เป็นส่วนหนึ่งของรายวิชา 32235265 เทคนิคการพยากรณ์และการ
ประยุกต์

(Forecasting Techniques and Application)

ภาคการศึกษาที่ 1 ปีการศึกษา 2567 สาขาวิทยาการข้อมูลและการวิเคราะห์

ข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

คำนำ

รายงานนี้เป็นส่วนหนึ่งของรายวิชา 32235265 เทคนิคการพยากรณ์และการประยุกต์ ทางคณะผู้จัดทำมีความประสงค์ที่จะศึกษา และค้นคว้าเกี่ยวกับการใช้เทคนิคการพยากรณ์และการประยุกต์ โดยการนำข้อมูลจำนวนผู้โดยสารที่ขึ้นเครื่องบินรายเดือนของสายการบิน United จากเว็บไซต์ kaggle เพื่อใช้ในการศึกษา จึงทำให้เกิดรายงานนี้ขึ้น

ทางคณะผู้จัดทำหวังว่ารายงานฉบับนี้จะสามารถสร้างประโยชน์ให้กับผู้ที่ต้องการศึกษา หรือเป็นแนวทางในการค้นคว้าและหาความรู้เพิ่มเติม ทั้งนี้หากรายงานฉบับนี้มีข้อผิดพลาดประการใด ทางคณะผู้จัดทำต้องขออภัยมา ณ ที่นี้

ทางคณะผู้จัดทำ

บทที่ 1

บทนำ

ผู้รับผิดชอบรายงาน : นางสาวปวีณา โชติประเสริฐ รหัสนักศึกษา 65030292

นายชัชดานัย รุ้ทัน รหัสนักศึกษา 65030737

สาขาวิทยาการข้อมูลและการวิเคราะห์ข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

1.1 ที่มาและความสำคัญ

ในปัจจุบันการเดินทางมีความสะดวกสบายและหลากหลายทางเลือกมากขึ้นเมื่อเทียบกับอดีต ไม่ว่าจะเป็น การเดิน การขับรถส่วนตัว หรือการใช้บริการรถโดยสารสาธารณะ ฯลฯ แต่หากต้องการเดินทางไกลในระยะเวลาอันสั้น โดยเฉพาะเมื่อมีธุระสำคัญหรือเหตุการณ์ฉุกเฉิน ทางเลือกที่ดีที่สุดมักจะเป็นการเดินทางโดยเครื่องบิน เนื่องจากสะดวกและประหยัดเวลาในการเดินทาง นอกจากนี้ การเปลี่ยนแปลงทางเศรษฐกิจและสังคมทำให้เกิดการติดต่อสื่อสารและการเชื่อมโยงระหว่างประเทศมากขึ้น ส่งผลให้การเดินทางทางอากาศได้รับความสนใจมากขึ้นเรื่อย ๆ จึงเกิดการแข่งขันสูงระหว่างสายการบินต่าง ๆ รวมถึงสายการบิน United ซึ่งเป็นหนึ่งในสายการบินหลักที่มีบริการทั่วโลก

ทางผู้จัดทำจึงเล็งเห็นว่าหากสามารถพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ได้ จะเป็นประโยชน์ต่อการวางแผนและเตรียมการล่วงหน้า เช่น การจัดสรรทรัพยากร การเพิ่มเที่ยวบิน หรือการจัดเตรียมบุคลากรที่เกี่ยวข้อง ทำให้สายการบินสามารถเตรียมพร้อมและสามารถคาดการณ์งบประมาณที่ต้องใช้ล่วงหน้าได้อย่างแม่นยำและเพิ่มความสามารถในการแข่งขันในตลาด ทางคณะผู้จัดทำจึงได้นำเทคนิคการพยากรณ์และการประยุกต์เข้ามาช่วยในการศึกษาครั้งนี้

1.2 วัตถุประสงค์ของการศึกษาการวิเคราะห์

1.2.1 เพื่อสร้างตัวแบบการพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United

1.2.2 เพื่อพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United

1.3 ขอบเขตของการศึกษาวิเคราะห์

ข้อมูลชุดนี้เป็นข้อมูลรายเดือนเกี่ยวกับจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน ตั้งแต่ช่วงเดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2553

1.3.1 วิเคราะห์ข้อมูลจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United และสร้างตัวแบบพยากรณ์

1.3.2 คัดเลือกตัวแบบที่เหมาะสม โดยใช้วิธีการปรับให้เรียบแบบเอ็กซ์โพเนนเชียลของ โฮลท์-วินเทอร์ (Holt-Winters Method) สำหรับตัวแบบเชิงคูณ และ พิจารณาจากค่า ACF และ PACF สำหรับตัวแบบ ARIMA

1.3.3 ตรวจสอบความเหมาะสมของตัวแบบด้วยค่าวัดความคลาดเคลื่อน MAPE (Mean Absolute Percentage Error) เมื่อเป็นตัวแบบเชิงคูณและตรวจสอบข้อสมมุติเมื่อเป็นตัวแบบ ARIMA

1.3.4 พยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ล่วงหน้า 12 เดือน

1.4 ผลที่คาดว่าจะได้รับการศึกษาวิเคราะห์

1.4.1 สามารถพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ในอนาคตจากตัวแบบพยากรณ์ที่เหมาะสม

1.4.2 เพื่อให้สายการบิน United ใช้ในการวางแผนและตัดสินใจในการเพิ่มหรือปรับลดเที่ยวบิน กำหนดราคาตั๋วเครื่องบิน รวมไปถึงการจัดสรรทรัพยากร เป็นต้น

บทที่ 2

วิธีการดำเนินการ

2.1 ข้อมูลที่ใช้ในการวิเคราะห์

ข้อมูลที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลรายเดือนเกี่ยวกับจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน ตั้งแต่ช่วงเดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2553 ซึ่งข้อมูลที่ได้มาจากการเก็บรวบรวมในเว็บไซต์ Kaggle โดย GABRIEL SANTELLO จากนั้นทางคณะผู้จัดทำได้นำข้อมูลมาทำการคัดกรองให้มีเฉพาะของสายการบิน United โดยข้อมูลที่ได้มีทั้งหมด 84 แถว 2 คอลัมน์ ซึ่งตัวอย่างข้อมูลมีลักษณะดังนี้

Date	Enplaned
01/2004	4434315
02/2004	4458657
03/2004	5302929
04/2004	5288871
05/2004	5408451
06/2004	5888419
07/2004	6114334
08/2004	6137271
09/2004	5070901
10/2004	5386126
11/2004	4838455
12/2004	5050794

ตารางที่ 2.1 ตัวอย่างตารางข้อมูลที่ใช้ในการวิเคราะห์

จากข้อมูลสามารถแบ่งลักษณะตัวแปรต่าง ๆ ที่ถูกนำมาใช้ในการวิเคราะห์ได้ดังนี้

ตัวแปรตาม : จำนวนผู้โดยสารที่ขึ้นเครื่องบินรายเดือน (Enplaned)

อนุกรมเวลา : วันที่ (Date) (เดือน/ปี)

2.2 เทคนิคการพยากรณ์

การวิเคราะห์ครั้งนี้จะนำเทคนิคการพยากรณ์อนุกรมเวลามาใช้ในการพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ซึ่งจะศึกษากราฟอนุกรมเวลาของข้อมูลที่มีการขึ้นลงหรือการเปลี่ยนแปลงของข้อมูลจำนวนผู้โดยสารที่ขึ้นเครื่องบินในระยะเวลา 6 ปี คือตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึง เดือนธันวาคม พ.ศ. 2553 โดยจะใช้การวิเคราะห์อนุกรมเวลา 2 วิธี คือ 1) วิธีการปรับให้เรียบแบบเอ็กซ์โพเนนเชียล ของ โฮลท์-วินเทอร์ (Holt-Winters Method) 2) วิธี Box-Jenkins ซึ่งมีหลักการดังนี้

2.2.1 วิธีการปรับให้เรียบแบบเอ็กซ์โพเนนเชียลของ โฮลท์-วินเทอร์ (Holt-Winters Method)

เป็นเทคนิคการพยากรณ์ที่จะมีการกำจัดอิทธิพลของความไม่แน่นอนออกไป เพื่อให้เรียบ และสามารถพยากรณ์หรือประมาณค่าตัวแปรในอนาคตได้ โดยการพยากรณ์จะให้ความสำคัญกับข้อมูลในอดีตซึ่งมีการกำหนดน้ำหนักที่จะนำมาใช้ในการคำนวณ

ในการวิเคราะห์นี้จะใช้วิธี Exponential Smoothing Adjusted for Trend and Seasonal Variation (Winters' Method) ในการพยากรณ์ โดยมีการกำหนดค่าคงที่สำหรับปรับให้เรียบ 3 ค่า คือ Alpha (α), Gamma (γ) และ Delta (δ) โดยที่

α = ค่าคงที่ทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์

γ = ค่าคงที่ทำให้เรียบระหว่างแนวโน้มจริงกับค่าประมาณของแนวโน้ม

δ = ค่าคงที่ทำให้เรียบระหว่างค่าฤดูกาลจริงกับค่าประมาณฤดูกาล

ตัวแบบที่ใช้ในการพยากรณ์

$$y_{t+k} = [b_{0,t} + b_{1,t}(k)] \times S_t + \epsilon_i$$

โดยที่

$$b_{0,t} = \alpha \left(\frac{y_t}{S_{t-L}} \right) + (1 - \alpha)(b_{0,t-1} + b_{1,t-1})$$

$$b_{1,t} = \gamma(b_{0,t} + b_{0,t-1})(1 - \gamma)b_{1,t-1}$$

$$\hat{S}_t = \delta \left(\frac{y_t}{S_{t-L}} \right) + (1 - \delta)(\hat{S}_{t-L})$$

และจะมีการกำหนดค่าเริ่มต้น b_0 , b_1 และ S_t เนื่องจากเราได้ทำการวิเคราะห์โดยใช้โปรแกรมสำเร็จรูป Minitab 18 จึงให้โปรแกรมกำหนดค่าเริ่มต้นให้

2.2.2 วิธี Box-Jenkins

เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยใช้ค่า Autocorrelation Function (ACF) และค่า Partial Autocorrelation Function (PACF) เป็นหลักในการพิจารณา และรูปแบบที่เลือกใช้จะอยู่ในกลุ่มของรูปแบบ SARIMA(P,D,Q)_s ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากค่าพยากรณ์ก่อนหน้า และความคลาดเคลื่อนของการพยากรณ์ก่อนหน้า โดยเป็นการผสมระหว่างตัวแบบ AR(p) และตัวแบบ MA(q)

ขั้นตอนในการวิเคราะห์อนุกรมเวลาโดยวิธี Box-Jenkins มี 6 ขั้นตอน คือ

- 1.การตรวจสอบความเป็นสมบัตินิ่งของข้อมูล (Stationary)
- 2.คำนวณหาฟังก์ชันอัตสหสัมพันธ์ Autocorrelation Function (ACF) และฟังก์ชันอัตสหสัมพันธ์บางส่วน Partial Autocorrelation Function (PACF)
3. การกำหนดตัวแบบ
4. การตรวจสอบความเหมาะสมของตัวแบบ
5. การประมาณค่าพารามิเตอร์
6. การนำตัวแบบไปใช้พยากรณ์

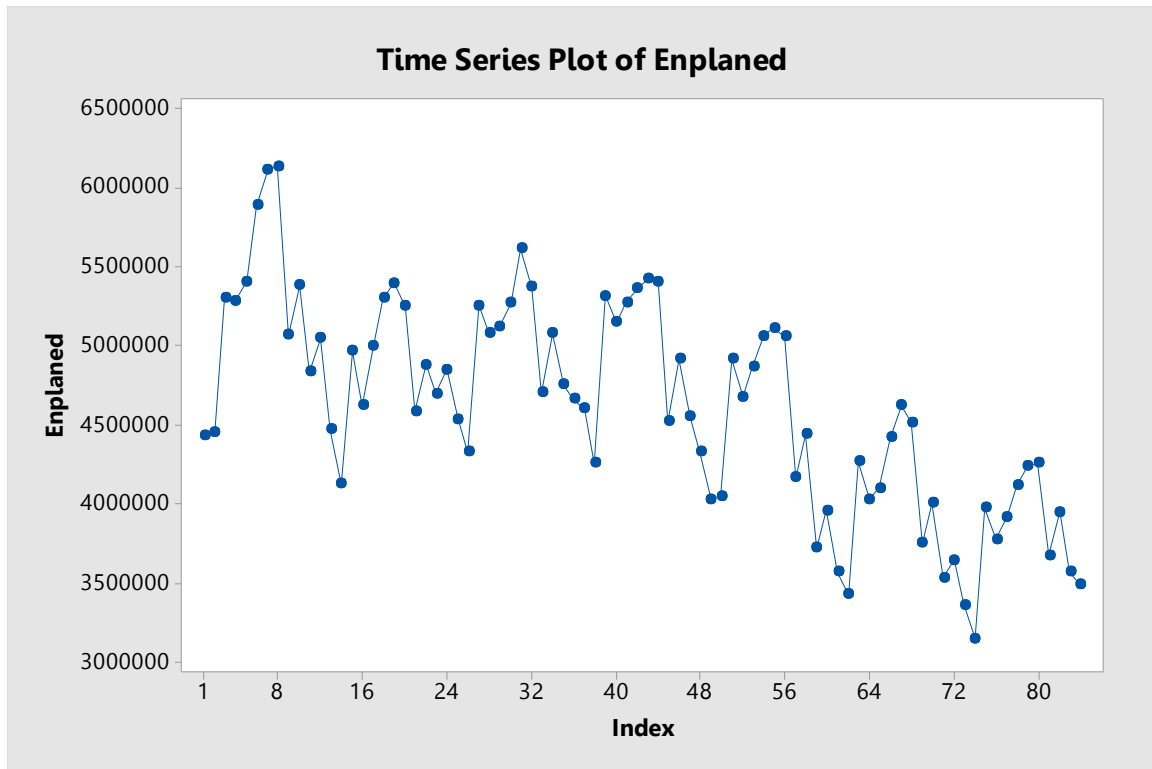
2.3 การคัดเลือกตัวแบบ

ในการพยากรณ์จะมีการพิจารณาสร้างตัวแบบพยากรณ์มากกว่าหนึ่งตัวแบบ ดังนั้นเพื่อคัดเลือกตัวแบบที่ทำให้ค่าพยากรณ์มีความแม่นยำมากที่สุด จะใช้การคำนวณค่าร้อยละของความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (Mean Absolute Percentage Error, MAPE) ของแต่ละตัวแบบพยากรณ์ที่เป็นตัวแบบเชิงคูณ และนำค่าที่ได้มาเปรียบเทียบกับกัน โดยเลือกตัวแบบพยากรณ์ที่ค่าร้อยละของความคลาดเคลื่อนเฉลี่ยสัมบูรณ์น้อยที่สุด ส่วนตัวแบบ ARIMA เราจะทำการตรวจสอบข้อสมมติด้วยกัน 3 ข้อ คือ 1. ตรวจสอบความเป็นอิสระกันของความคลาดเคลื่อน 2. ตรวจสอบการแจกแจงปกติของตัวแบบ 3. ตรวจสอบความคงที่ของความแปรปรวนของค่าคลาดเคลื่อน

บทที่ 3

ขั้นตอนการดำเนินการวิเคราะห์ข้อมูล

3.1 พิจารณาความสัมพันธ์ระหว่างตัวแปรตามและอนุกรมเวลา



รูปที่ 3.1 กราฟอนุกรมเวลาของจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ตั้งแต่เดือน มกราคม 2547 ถึง ธันวาคม 2553

กราฟอนุกรมเวลาข้างต้น สังเกตได้ว่าจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United มีค่าต่ำสุดในช่วงเดือนกุมภาพันธ์ จากนั้นจำนวนผู้โดยสารที่ขึ้นเครื่องบินจะเพิ่มขึ้นเรื่อย ๆ จนสูงสุดในช่วงเดือนกรกฎาคม และสิงหาคม หลังจากนั้นจำนวนผู้โดยสารที่ขึ้นเครื่องบินจะลดลงเรื่อย ๆ จนถึงเดือนกุมภาพันธ์ ซึ่งจากกราฟจะเห็นได้ว่าจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United มีรูปแบบการขึ้นลงในลักษณะเดิมทุก ๆ ปี แสดงว่าจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ได้รับอิทธิพลจากฤดูกาล เมื่อทำการสังเกต พบว่า จุดต่ำสุดและจุดสูงสุดของฤดูกาลแต่ละช่วงไม่ขนานกัน จึงทำการเลือกใช้ตัวแบบเชิงคูณ

3.2 การพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียลของ โฮลท์-วินเทอร์ (Holt-Winters Method)

เนื่องจากข้อมูลได้รับอิทธิพลจากฤดูกาล และจุดต่ำสุดและจุดสูงสุดของฤดูกาลแต่ละช่วงไม่ยาวนานกัน ดังนั้น จึงเลือกวิธี Exponential Smoothing Adjusted for Trend and Seasonal Variation (Winters' Method) ในการพยากรณ์

การวิเคราะห์โดยใช้ตัวแบบฤดูกาลเชิงคูณ ผู้วิเคราะห์ได้กำหนดค่าคงที่ปรับให้เรียบ 3 ค่า คือ Alpha (α) Gamma (γ) และ Delta (δ) โดยจะกำหนด α , γ และ δ ให้มีค่าระหว่าง 0.1 – 0.3 เพื่อให้ข้อมูลและค่าประมาณแนวโน้มถูกปรับให้เรียบมาก และกำหนดค่าคงที่ปรับเรียบให้มีค่าสูงระหว่าง 0.7 - 0.8 เพื่อให้ข้อมูลและค่าประมาณแนวโน้มมีความผันแปรขึ้นลงเป็นไปตามปัจจุบันมากที่สุด จากการปรับให้เรียบของข้อมูลด้วยค่าคงที่ปรับให้เรียบต่าง ๆ กัน จะได้ค่าความคลาดเคลื่อนเฉลี่ยแบบสัมบูรณ์ (MAPE) เพื่อนำมาเปรียบเทียบกันแล้วเลือกค่าคงที่ปรับให้เรียบที่ให้ค่าความคลาดเคลื่อนเฉลี่ยแบบสัมบูรณ์ (MAPE) ที่ต่ำที่สุด เพราะฉะนั้น จะสามารถกำหนดค่า Alpha (α), Gamma (γ) และ Delta (δ) ที่มีค่าระหว่าง 0.1 - 0.3 และ 0.7 – 0.8 ได้ทั้งหมด 125 แบบ ซึ่งเราได้คำนวณโดยใช้โปรแกรม Minitab 18 ได้ค่าดังตารางต่อไปนี้

ค่าคงที่ปรับเรียบ			ค่าความคลาดเคลื่อนเฉลี่ยแบบสัมบูรณ์ (MAPE)
Alpha (α)	Gamma (γ)	Delta (δ)	
0.1	0.1	0.1	4.01321
0.1	0.1	0.2	4.07650
0.1	0.1	0.3	4.14067
0.1	0.1	0.7	4.30152
0.1	0.1	0.8	4.35293
0.1	0.2	0.1	4.50623
0.1	0.2	0.2	4.42391
0.1	0.2	0.3	4.36203
0.1	0.2	0.7	4.32042
0.1	0.2	0.8	4.40944
0.1	0.3	0.1	4.03904
0.1	0.3	0.2	4.02702
0.1	0.3	0.3	4.02643
0.1	0.3	0.7	4.46912
0.1	0.3	0.8	4.67275

0.1	0.7	0.1	4.22107
0.1	0.7	0.2	4.52179
0.1	0.7	0.3	4.87004
0.1	0.7	0.7	7.42490
0.1	0.7	0.8	9.59637
0.1	0.8	0.1	4.39376
0.1	0.8	0.2	4.80445
0.1	0.8	0.3	5.27011
0.1	0.8	0.7	10.5383
0.1	0.8	0.8	14.0023
0.2	0.1	0.1	3.07843
0.2	0.1	0.2	3.17178
0.2	0.1	0.3	3.28303
0.2	0.1	0.7	3.88735
0.2	0.1	0.8	4.11631
0.2	0.2	0.1	3.06494
0.2	0.2	0.2	3.20664
0.2	0.2	0.3	3.37776
0.2	0.2	0.7	4.43306
0.2	0.2	0.8	4.78726
0.2	0.3	0.1	3.21403
0.2	0.3	0.2	3.41206
0.2	0.3	0.3	3.69543
0.2	0.3	0.7	5.35714
0.2	0.3	0.8	5.96954
0.2	0.7	0.1	4.33054
0.2	0.7	0.2	5.05394
0.2	0.7	0.3	6.10260
0.2	0.7	0.7	13.5202
0.2	0.7	0.8	16.4006
0.2	0.8	0.1	4.41829
0.2	0.8	0.2	5.26188

0.2	0.8	0.3	6.53079
0.2	0.8	0.7	15.0589
0.2	0.8	0.8	18.3299
0.3	0.1	0.1	2.64523
0.3	0.1	0.2	2.73087
0.3	0.1	0.3	2.83918
0.3	0.1	0.7	3.51066
0.3	0.1	0.8	3.73729
0.3	0.2	0.1	2.78996
0.3	0.2	0.2	2.90131
0.3	0.2	0.3	3.05338
0.3	0.2	0.7	4.16442
0.3	0.2	0.8	4.53888
0.3	0.3	0.1	2.98305
0.3	0.3	0.2	3.15534
0.3	0.3	0.3	3.39432
0.3	0.3	0.7	5.04416
0.3	0.3	0.8	5.66175
0.3	0.7	0.1	2.97509
0.3	0.7	0.2	3.11274
0.3	0.7	0.3	3.40146
0.3	0.7	0.7	6.77301
0.3	0.7	0.8	8.03377
0.3	0.8	0.1	2.81304
0.3	0.8	0.2	2.90762
0.3	0.8	0.3	3.13380
0.3	0.8	0.7	6.12691
0.3	0.8	0.8	7.32650
0.7	0.1	0.1	2.26166
0.7	0.1	0.2	2.26955
0.7	0.1	0.3	2.28123
0.7	0.1	0.7	2.34615

0.7	0.1	0.8	2.37587
0.7	0.2	0.1	2.30962
0.7	0.2	0.2	2.31562
0.7	0.2	0.3	2.32178
0.7	0.2	0.7	2.36560
0.7	0.2	0.8	2.38945
0.7	0.3	0.1	2.31239
0.7	0.3	0.2	2.31639
0.7	0.3	0.3	2.32232
0.7	0.3	0.7	2.37963
0.7	0.3	0.8	2.40990
0.7	0.7	0.1	2.45750
0.7	0.7	0.2	2.44820
0.7	0.7	0.3	2.44650
0.7	0.7	0.7	2.55415
0.7	0.7	0.8	2.60171
0.7	0.8	0.1	2.48733
0.7	0.8	0.2	2.47996
0.7	0.8	0.3	2.48653
0.7	0.8	0.7	2.63297
0.7	0.8	0.8	2.69419
0.8	0.1	0.1	2.25070
0.8	0.1	0.2	2.25551
0.8	0.1	0.3	2.26009
0.8	0.1	0.7	2.29057
0.8	0.1	0.8	2.30297
0.8	0.2	0.1	2.27801
0.8	0.2	0.2	2.28223
0.8	0.2	0.3	2.28836
0.8	0.2	0.7	2.32177
0.8	0.2	0.8	2.33455
0.8	0.3	0.1	2.31838

0.8	0.3	0.2	2.32730
0.8	0.3	0.3	2.33591
0.8	0.3	0.7	2.38254
0.8	0.3	0.8	2.39615
0.8	0.7	0.1	2.50998
0.8	0.7	0.2	2.52121
0.8	0.7	0.3	2.53190
0.8	0.7	0.7	2.60198
0.8	0.7	0.8	2.62755
0.8	0.8	0.1	2.57891
0.8	0.8	0.2	2.59287
0.8	0.8	0.3	2.60592
0.8	0.8	0.7	2.69025
0.8	0.8	0.8	2.71848

ตารางที่ 3.2 แสดงค่าความคลาดเคลื่อนเฉลี่ยแบบสัมบูรณ์ และค่าคงที่ปรับให้เรียบ

ของจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United

จากตารางข้างต้น Alpha (α), Gamma (γ) และ Delta (δ) ที่มีค่า 0.8, 0.1 และ 0.1 ตามลำดับ ให้ค่าความคลาดเคลื่อนเฉลี่ยแบบสัมบูรณ์ (MAPE) ต่ำสุด คือ 2.25070

ตัวแบบที่ใช้ในการพยากรณ์ คือ

$$y_{t+k} = [b_{0,t} + b_{1,t}(k)] \times S_t + \epsilon_i$$

$$\text{โดยที่ } b_{0,t} = 0.8\left(\frac{y_t}{S_{t-L}}\right) + 0.2(b_{0,t-1} + b_{1,t-1})$$

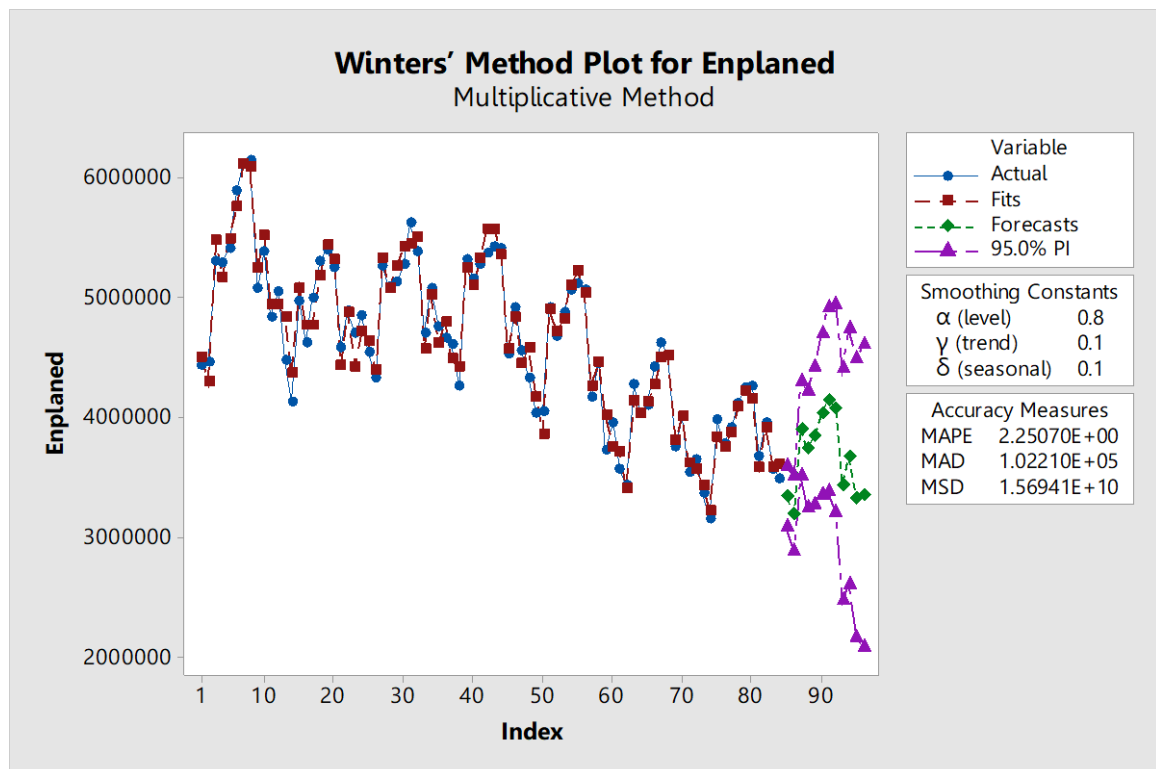
$$b_{1,t} = 0.1(b_{0,t} + b_{0,t-1}) + 0.9b_{1,t-1}$$

$$\hat{S}_t = 0.1\left(\frac{y_t}{S_{t-L}}\right) + 0.9\hat{S}_{t-L}$$

เมื่อ K คือ หน่วยเวลาของการพยากรณ์ล่วงหน้า

L คือ คาบเวลาในการเกิดฤดูกาล = 12

S_1, S_2, \dots, S_{12} คือ ค่าดัชนีฤดูกาล ตั้งแต่ มกราคม ถึง ธันวาคม ตามลำดับ



รูปที่ 3.3 กราฟแสดงวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียลของ โฮลท์-วินเทอร์
ของจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United

กราฟที่ได้จากโปรแกรม Minitab 18 รูปที่ 3.3 เส้นสีน้ำเงินแทนค่าจริงของข้อมูล เส้นสีแดงแทนค่าของการพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินในอดีต เส้นสีเขียวแทนค่าพยากรณ์ของจำนวนผู้โดยสารที่ขึ้นเครื่องบินในอนาคต และเส้นสีม่วงแทนขอบเขตของค่าพยากรณ์ของจำนวนผู้โดยสารที่ขึ้นเครื่องบินขอบเขตล่างและขอบเขตบน โดยเราได้ทำการพยากรณ์ค่าในอนาคตออกมาทั้งหมด 12 ค่า ดังนี้

เดือน/ปี	ค่าพยากรณ์
มกราคม 2554	3341672
กุมภาพันธ์ 2554	3203611
มีนาคม 2554	3910089
เมษายน 2554	3739791
พฤษภาคม 2554	3851438
มิถุนายน 2554	4035949
กรกฎาคม 2554	4149783
สิงหาคม 2554	4077677
กันยายน 2554	3441364
ตุลาคม 2554	3676391

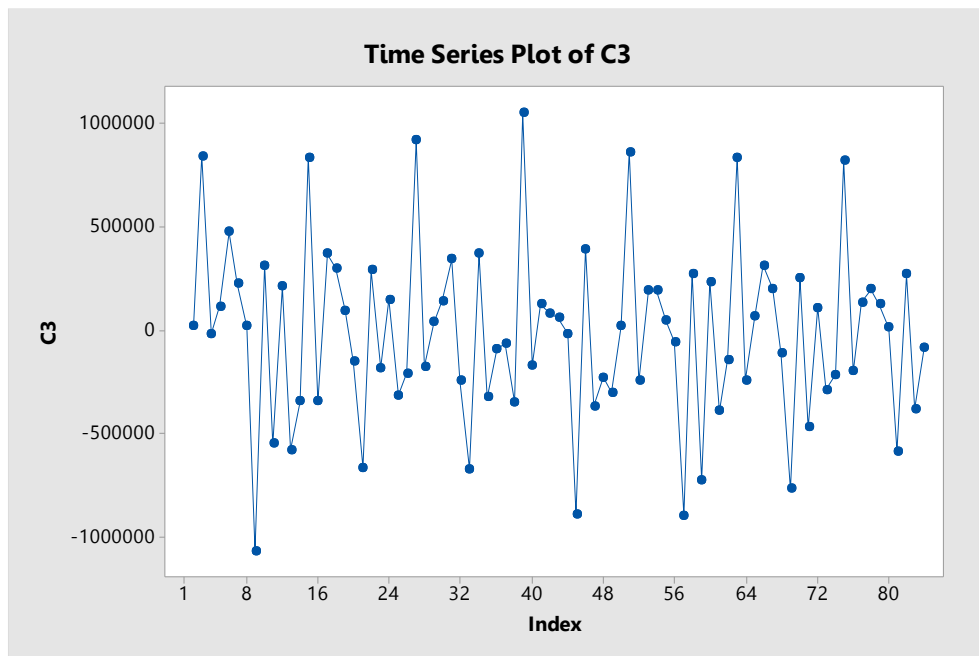
พฤศจิกายน 2554	3327304
ธันวาคม 2554	3353457

ตารางที่ 3.3 แสดงค่าพยากรณ์ 12 เดือน ด้วยตัวแบบเชิงคูณ

3.3 วิธี Box-Jenkins

3.3.1 ตรวจสอบความเป็นสมบัตินิ่งของข้อมูล (Stationary)

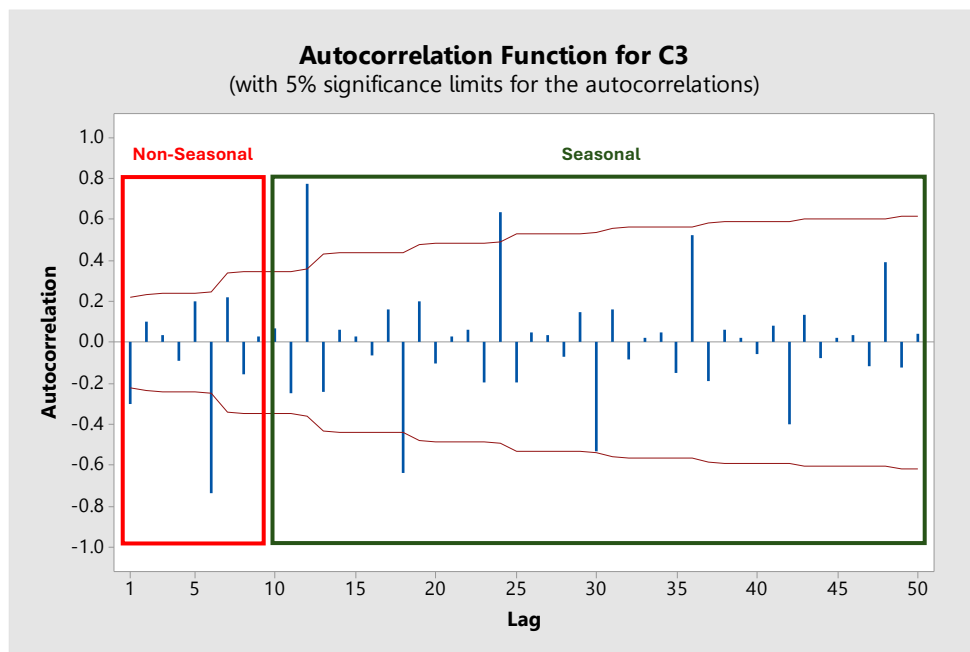
จากกราฟรูปที่ 3.1 สามารถสังเกตได้ว่า ข้อมูลอนุกรมชุดนี้ไม่มีความ Stationary จึงจะต้องทำการ Differences เพื่อปรับให้ค่าของข้อมูลอนุกรมเวลามีความ Stationary



รูปที่ 3.4 กราฟอนุกรมเวลาของจำนวนผู้โดยสารที่ขึ้นเครื่องบิน เมื่อทำการ Differences

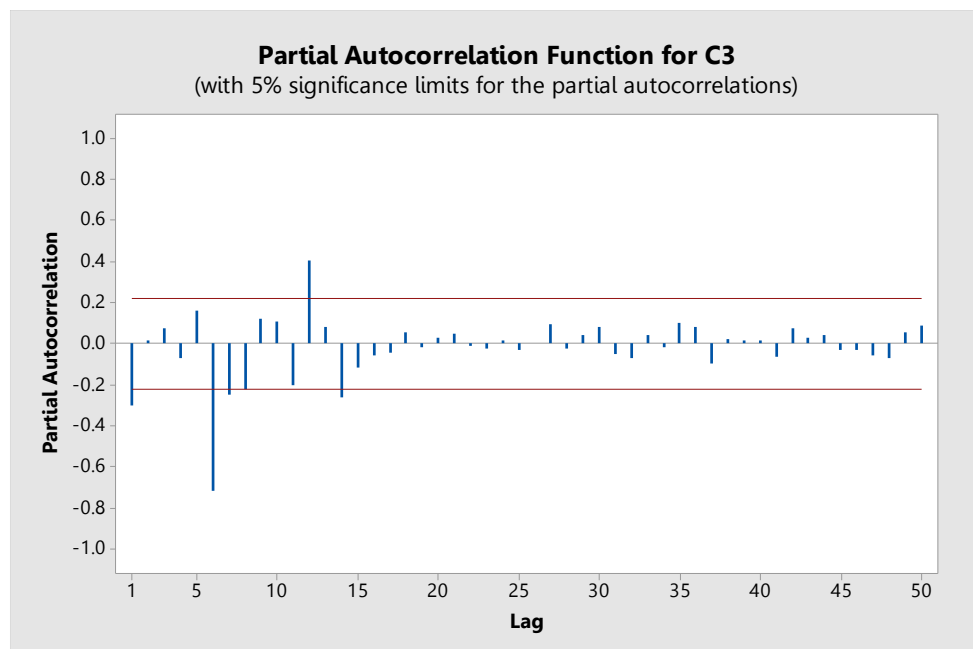
กราฟอนุกรมเวลาที่ทำการ Differences แล้ว พบว่าข้อมูลอนุกรมเวลามีความ Stationary จึงได้นำข้อมูลไปใช้ในการวิเคราะห์ขั้นตอนถัดไป

3.3.2 คำนวณหาฟังก์ชันอัตสหสัมพันธ์ (Auto correlation function : ACF) และ ฟังก์ชันอัตสหสัมพันธ์บางส่วน (Partial auto correlation function : PACF) ของข้อมูลพยากรณ์



รูปที่ 3.5 กราฟฟังก์ชันอัตสหสัมพันธ์ (Auto correlation function : ACF)

จากกราฟ เมื่อทำการสังเกต พบว่า lag บนกราฟ ACF ในส่วนของ Non-Seasonal และ Seasonal มีลักษณะลดลงสู่ศูนย์ในลักษณะ tail off



รูปที่ 3.6 กราฟฟังก์ชันอัตสหสัมพันธ์บางส่วน (Partial auto correlation function : PACF)

จากกราฟ เมื่อทำการสังเกต พบว่า lag บนกราฟ PACF ในส่วนของ Non-Seasonal และ Seasonal มีลักษณะลดลงแบบรวดเร็ว ซึ่งจะมีลักษณะ cuts off หลัง lag 1 ทั้งในส่วนของ Non-Seasonal และ Seasonal ดังนั้นตัวแบบที่เหมาะสมสำหรับใช้พยากรณ์ข้อมูลชุดนี้ควรเป็นตัวแบบ $ARIMA(1,1,0) \times (1,0,0)_{12}$

3.3.3 กำหนดตัวแบบ (Model Identification)

จาก $ARIMA(1,1,0) \times (1,0,0)_{12}$ จะสามารถเขียนตัวแบบได้ดังนี้

โดยที่ตัวแบบอัตราการถดถอยอันดับที่ P ของ $ARIMA(1,1,0)$ คือ

$$W_t = \delta + \phi_1 W_{t-1} + u_t$$

เมื่อ

$$W_t = X_t - X_{t-1}$$

จะได้ว่า

$$X_t - X_{t-1} = \delta + \phi_1 (X_{t-1} - X_{t-2}) + u_t$$

$$X_t = \delta + \phi_1 (X_{t-1} - X_{t-2}) + u_t + X_{t-1}$$

และ ตัวแบบอัตราการถดถอยแบบฤดูกาลอันดับที่ P ของ $ARIMA(1,0,0)_{12}$ คือ

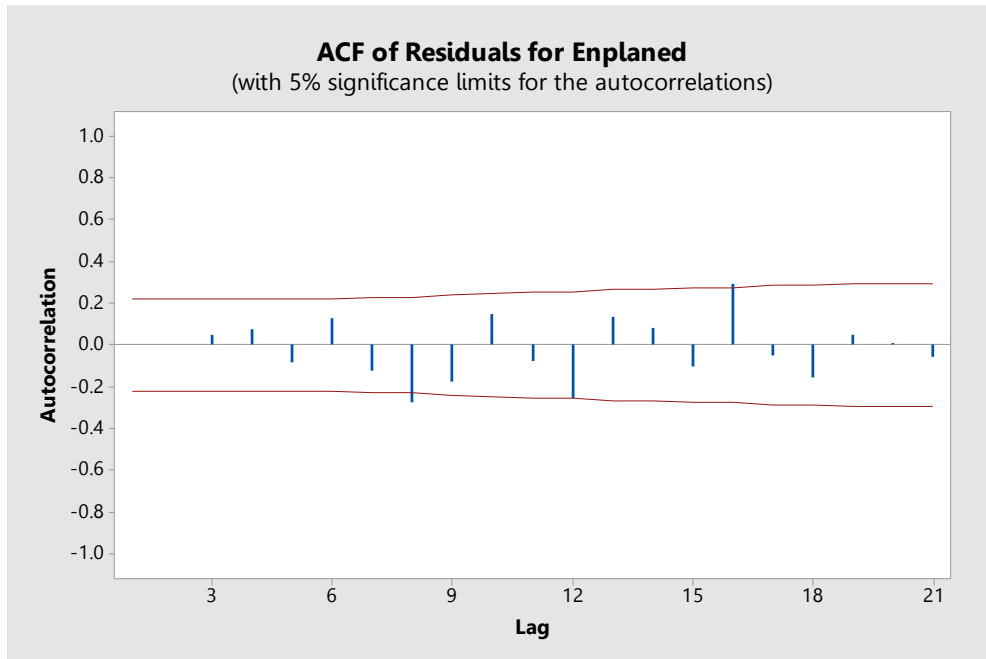
$$X_t = \delta + \Phi_1 (X_{t-1}) + u_t$$

ดังนั้นตัวแบบของ $ARIMA(1,1,0) \times (1,0,0)_{12}$ คือ

$$X_t = \delta + \phi_1 (X_{t-1} - X_{t-2}) + \Phi_1 (X_{t-1}) + u_t + X_{t-1}$$

3.3.4 ตรวจสอบข้อสมมุติของตัวแบบ

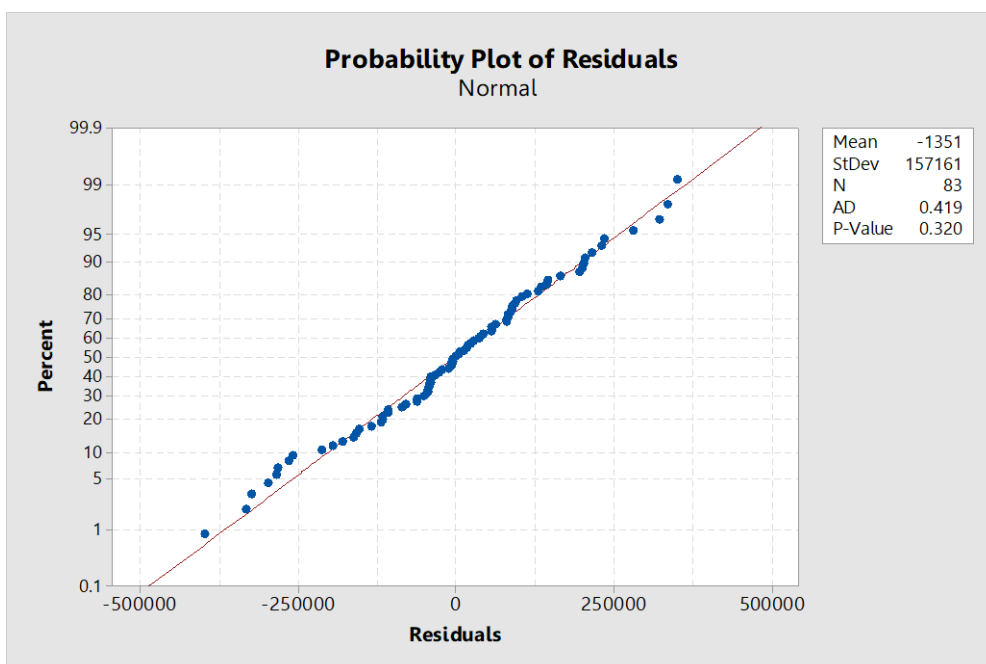
1. ตรวจสอบความเป็นอิสระกันของความคาดเคลื่อน



รูปที่ 3.7 แผนภาพฟังก์ชันอัตโนมัติสัมพันธ์ของเศษเหลือสำหรับจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United

จากแผนภาพจะเห็นได้ว่า ไม่มีลักษณะ cuts off ของ lag สำหรับเศษเหลือใดเลยจึงสามารถสรุปได้ว่าตัวแบบนี้มีความอิสระกันของความคาดเคลื่อน

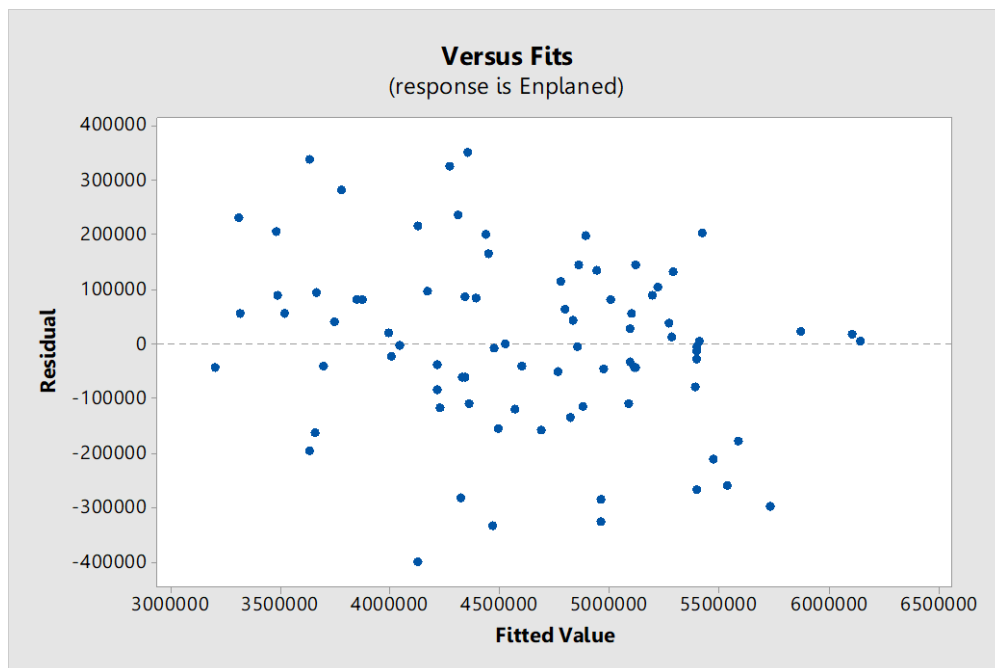
2. ตรวจสอบการแจกแจงปกติของตัวแบบ



รูปที่ 3.8 แผนภาพความน่าจะเป็นปกติของส่วนเหลือ

จากแผนภาพความน่าจะเป็นปกติของส่วนเหลือ จะเห็นได้ว่าการกระจายของจุดในแผนภาพนั้น ใกล้เคียงกับเส้นตรง แสดงว่ามีการแจกแจงใกล้เคียงกับการแจกแจงปกติ และเมื่อทำการทดสอบ Anderson-darling พบว่า $P\text{-Value} = 0.320 > 0.05$ จึงยอมรับสมมติฐานหลักนั้นคือความคาดเคลื่อนมีการแจกแจงแบบปกติ

3. ตรวจสอบความคงที่ของความแปรปรวนของค่าคลาดเคลื่อน



รูปที่ 3.9 แผนภาพการกระจายระหว่างส่วนเหลือและค่าประมาณของจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United

จากแผนภาพจะแสดงให้เห็นว่าจุดบนแผนภาพมีการกระจายแบบสุ่มอยู่ภายในรอบ ๆ เส้นศูนย์กลางของภาพ จึงไม่เป็นการผิดข้อสมมุตินี้ ดังนั้นจึงสามารถสรุปได้ว่า ตัวแบบนี้มีความแปรปรวนของค่าคลาดเคลื่อนคงที่

3.3.5 การประมาณค่าพารามิเตอร์ (Parameter Estimation)

จากตัวแบบ $ARIMA(1,1,0) \times (1,0,0)_{12}$ จะสามารถเขียนตัวแบบได้ดังนี้

Type	Coef	P-Value
AR 1	0.106	0.002
SAR 12	0.0406	0.000

ตารางที่ 3.4 แสดงผลการวิเคราะห์ ARIMA

$$X_t = \delta + \phi_1 (X_{t-1} - X_{t-2}) + \Phi_1 (X_{t-1}) + u_t + X_{t-1}$$

ดังนั้นสมการพยากรณ์ คือ

$$\hat{X}_t = (-0.336)(X_{t-1} - X_{t-2}) + (0.9785) (X_{t-1}) + X_{t-1}$$

จะได้ว่า

$$\hat{X}_t = (0.664X_{t-1}) - (0.336X_{t-2}) + (0.9785) (X_{t-1})$$

3.3.5 การนำไปใช้ในการพยากรณ์ (Forecasting)

จากสมการพยากรณ์เราได้ทำการพยากรณ์ค่าในอนาคตออกมาทั้งหมด 12 ค่า ดังนี้

เดือน/ปี	ค่าพยากรณ์
มกราคม 2554	4316426
กุมภาพันธ์ 2554	3964364
มีนาคม 2554	5001239
เมษายน 2554	4837481
พฤษภาคม 2554	4962770
มิถุนายน 2554	5045575
กรกฎาคม 2554	5105368
สิงหาคม 2554	5092473
กันยายน 2554	4224762
ตุลาคม 2554	4612991
พฤศจิกายน 2554	4254554
ธันวาคม 2554	4035023

ตารางที่ 3.5 แสดงค่าพยากรณ์ 12 เดือน ด้วยตัวแบบ ARIMA

บทที่ 4

สรุปผลและอภิปรายผลการศึกษา

สรุปผลการศึกษา

การศึกษานี้ได้ทำการศึกษาเรื่องการพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United โดยข้อมูลที่ใช้ในการพยากรณ์ได้จากข้อมูลรายเดือนเกี่ยวกับจำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน ตั้งแต่ช่วงเดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2553 ซึ่งข้อมูลที่ได้มาจากการเก็บรวบรวมใน เว็บไซต์ Kaggle โดย GABRIEL SANTELLO และได้นำเทคนิคมาใช้ในการพยากรณ์ 2 วิธีคือ

- 1) วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียลของ โฮลท์-วินเทอร์ (Holt-Winters Method)
- 2) วิธี Box-Jenkins

จากผลการวิเคราะห์สามารถสรุปตัวแบบที่เหมาะสมของทั้ง 2 วิธี สำหรับการพยากรณ์จำนวนผู้โดยสารที่ขึ้นเครื่องบินของสายการบิน United ดังนี้

วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล ของ โฮลท์-วินเทอร์ (Holt-Winters Method)

ตัวแบบที่ใช้ในการพยากรณ์ มีค่า Alpha (α), Gamma (γ) และ Delta (δ) ที่มีค่า 0.8, 0.1 และ 0.1 ตามลำดับ ซึ่งให้ค่าความคลาดเคลื่อนเฉลี่ยแบบสัมบูรณ์ (MAPE) ต่ำสุด คือ 2.25070

วิธี Box-Jenkins

ตัวแบบ ARIMA ที่เลือกใช้สำหรับข้อมูลชุดนี้คือ $ARIMA(1,1,0)(1,0,0)_{12}$ เนื่องจากเมื่อทำการตรวจสอบข้อสมมุติแล้วพบว่า ไม่มีการผิดข้อสมมุติ และเป็นการเลือกใช้ตัวแบบ ARIMA ที่มีอันดับ p, d, q ที่น้อยที่สุด