บทที่ 3

การวิเคราะห์ข้อมูล

การวิจัยเรื่องการเปรียบเทียบวิธีการพยากรณ์ความต้องการใช้ไฟฟ้าในส่วนภูมิภาคนี้ เป็นการศึกษาวิเคราะห์ข้อมูลเพื่อการพยากรณ์ข้อมูลปริมาณความต้องการใช้ไฟฟ้าในแต่ละประเภท ซึ่งใช้วิธีการวิเคราะห์อนุกรมเวลา วิธีการวิเคราะห์การถดถอย และวิธีการพยากรณ์ของการไฟฟ้าส่วนภูมิภาค โดยการวิเคราะห์อนุกรมเวลาประกอบด้วย วิธีการบอกซ์ – เจนกินส์ วิธีการปรับให้เรียบแบบเลขชี้กำลัง และวิธีการแยกองค์ประกอบ ผู้วิจัยจะศึกษาผลการเปรียบเทียบวิธีการพยากรณ์ทั้ง 5 วิธี จากการเปรียบเทียบค่าความคลาดเคลื่อนของการพยากรณ์ โดยใช้ค่าเฉลี่ยของเปอร์เซ็นต์ค่าความคลาดเคลื่อนสัมบูรณ์ (MAPE) เป็นค่าที่ใช้พิจารณาว่าวิธีการพยากรณ์วิธีใดให้ค่า MAPE ต่ำสุด วิธีนั้นจะเป็นวิธีที่เหมาะสมกับข้อมูลของแต่ละประเภทนั้น ๆ ซึ่งจะได้ตัวแบบที่เหมาะสมเพื่อนำไปใช้พยากรณ์ค่าต่อไป ดังนั้นจะเห็นได้ว่าขั้นตอนของระบบงานพยากรณ์นั้น แบ่งออกเป็น 2 ขั้นตอนใหญ่ ๆ คือ ขั้นตอนของการสร้างตัวแบบพยากรณ์ และขั้นตอนของการพยากรณ์ โดยมีรายละเอียดของขั้นตอนต่าง ๆ ในระบบงานพยากรณ์ตามแผนภูมิแผนภาพที่ 3.1 ซึ่งลักษณะของข้อมูลและวิธีการพยากรณ์ ที่นำมาใช้ในการวิเคราะห์ข้อมูลครั้งนี้ มีรายละเอียดดังต่อไปนี้

3.1 การวิเคราะห์ข้อมูล

การศึกษาวิเคราะห์ข้อมูลที่ใช้ในการวิจัยครั้งนี้ เป็นการพยากรณ์ตัวแบบของปริมาณความต้องการใช้ไฟฟ้าในแต่ละประเภทของกลุ่มผู้ใช้ไฟฟ้า ซึ่งแบ่งออกเป็น 5 กลุ่ม โดยจะจำแนกตามรายภาคต่าง ๆ ของเขตการไฟฟ้าส่วนภูมิภาค ข้อมูลของปริมาณความต้องการใช้ไฟฟ้า ที่นำมาวิเคราะห์เป็นอนุกรมเวลาที่ประกอบด้วยข้อมูลทั้งหมด 20 ประเภท โดยการเก็บรวบรวมข้อมูลของแผนกสถิติการใช้ไฟฟ้า กองเศรษฐกิจพลังไฟฟ้า การไฟฟ้าส่วนภูมิภาค ซึ่งประกอบด้วยข้อมูลดังต่อไปนี้

1. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทบ้านอยู่อาศัยขนาดเล็กและบ้านอยู่อาศัยขนาดใหญ่ ในภาคเหนือ

2. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทบ้านอยู่อาศัยขนาดเล็กและบ้านอยู่อาศัยขนาดใหญ่ ในภาคตะวันออกเฉียงเหนือ

3. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทบ้านอยู่อาศัยขนาดเล็กและบ้านอยู่อาศัยขนาดใหญ่ ในภาคกลาง

4. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทบ้านอยู่อาศัยขนาดเล็ก และบ้านอยู่อาศัยขนาดใหญ่ ในภาคใต้

5. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการขนาดเล็ก กิจการขนาดกลาง และกิจการขนาดใหญ่ ในภาคเหนือ

6. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการขนาดเล็ก กิจการขนาดกลาง และกิจการขนาดใหญ่ ในภาคตะวันออกเฉียงเหนือ

7. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการขนาดเล็ก กิจการขนาดกลาง และกิจการขนาดใหญ่ ในภาคกลาง

8. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการขนาดเล็ก กิจการขนาดกลาง และกิจการขนาดใหญ่ ในภาคใต้

9. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการเฉพาะอย่าง ในภาคเหนือ

10. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการเฉพาะอย่าง ในภาคตะวันออกเฉียงเหนือ

11. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการเฉพาะอย่าง ในภาคกลาง

12. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทกิจการเฉพาะอย่าง ในภาคใต้

13. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทส่วนราชการและองค์กรที่ไม่แสวงหากำไร ในภาคเหนือ

14. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทส่วนราชการและองค์กรที่ไม่แสวงหากำไร ในภาคตะวันออกเฉียงเหนือ

15. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทส่วนราชการและองค์กรที่ไม่แสวงหากำไร ในภาคกลาง

16. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทส่วนราชการและองค์กรที่ไม่แสวงหากำไร ในภาคใต้

17. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทสูบน้ำเพื่อการเกษตรและไฟชั่วคราว ในภาคเหนือ

18. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทสูบน้ำเพื่อการเกษตรและไฟชั่วคราว ในภาคตะวันออกเฉียงเหนือ

19. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทสูบน้ำเพื่อการเกษตรและไฟชั่วคราว ในภาคกลาง

20. ข้อมูลปริมาณความต้องการใช้ไฟฟ้าของกลุ่มผู้ใช้ไฟฟ้าประเภทสูบน้ำเพื่อการเกษตรและไฟชั่วคราว ในภาคใต้

ในการวิเคราะห์ข้อมูลของปริมาณความต้องการใช้ไฟฟ้าแต่ละประเภท โดยวิธีการวิเคราะห์อนุกรมเวลา และวิธีการวิเคราะห์การถดถอย ซึ่งจะพิจารณาวิเคราะห์แต่ละตัวแบบดังนี้

1. ตัวแบบอนุกรมเวลา

ข้อมูลที่นำมาวิเคราะห์ด้วยวิธีการวิเคราะห์อนุกรมเวลา เป็นข้อมูลรายเดือน ระหว่างปีงบประมาณ 2536 – 2542 เริ่มตั้งแต่เดือนตุลาคม ปี พ.ศ. 2535 ถึงเดือนกันยายน ปี พ.ศ. 2542 ขนาดของข้อมูลทั้งหมดมีจำนวน 84 ค่า โดยจะใช้ข้อมูลทั้งหมดนี้ ในการวิเคราะห์หาตัวแบบอนุกรมเวลาที่เหมาะสมในเชิงสถิติ สำหรับใช้ในการพยากรณ์ค่า และนำค่าพยากรณ์ทั้งหมดที่ได้จากตัวแบบดังกล่าวที่เป็นรายเดือน มาหาค่าผลรวมรายปีของแต่ละปีงบประมาณ เพื่อมาเปรียบเทียบกับข้อมูลจริงที่เป็นค่าผลรวมรายปี ตั้งแต่ปี 2536 ถึงปี 2542 โดยพิจารณาจากค่า MAPE เพื่อใช้ในการตรวจสอบหาวิธีการพยากรณ์ที่เหมาะสม

2. ตัวแบบการถดถอย

ข้อมูลที่นำมาวิเคราะห์ด้วยวิธีการวิเคราะห์การถดถอย เป็นข้อมูลรายปี โดยเริ่มตั้งแต่ปีงบประมาณ 2524 – 2542 ขนาดของข้อมูลทั้งหมดมีจำนวน 19 ค่า โดยจะใช้ข้อมูลทั้งหมดนี้ ในการวิเคราะห์หาตัวแบบการถดถอยที่เหมาะสมในเชิงสถิติ สำหรับใช้ในการพยากรณ์ค่า และนำตัวแบบที่ได้นั้นมาหาค่าพยากรณ์รายปี เพื่อมาเปรียบเทียบกับข้อมูลจริงรายปี ตั้งแต่ปี 2536 ถึงปี 2542 โดยพิจารณาจากค่า MAPE เพื่อใช้ในการตรวจสอบหาวิธีการพยากรณ์ที่เหมาะสม

ตัวแบบหรือวิธีการพยากรณ์วิธีใดให้ค่า MAPE ต่ำสุด วิธีนั้นจะเป็นวิธีที่เหมาะสมกับข้อมูลของแต่ละชุดนั้น ๆ ซึ่งจะได้ตัวแบบที่เหมาะสมที่ใช้ในการพยากรณ์ค่าในอนาคต โดยจะใช้ตัวแบบดังกล่าวนั้นในการพยากรณ์ค่าในปี 2543 ถึงปี 2545

วิธีการทางสถิติในการวิเคราะห์ข้อมูลครั้งนี้ ใช้การวิเคราะห์อนุกรมเวลา และการวิเคราะห์การถดถอย โดยวิธีพยากรณ์ที่ใช้ในการวิเคราะห์อนุกรมเวลา ประกอบด้วยวิธีการบอกซ์ – เจนกินส์ วิธีการปรับให้เรียบแบบเลขชี้กำลัง และวิธีการแยกองค์ประกอบ มีวิธีการดังนี้

3.1.1 วิธีการบอกซ์ – เจนกินส์

วิธีการพยากรณ์อนุกรมเวลาโดยวิธีการบอกซ์ – เจนกินส์นี้ ใช้ได้กับข้อมูลอนุกรมเวลาที่มีองค์ประกอบใด ๆ เช่น แนวโน้ม และวัฏจักร หรือฤดูกาล ซึ่งอาจจะมีองค์ประกอบมากกว่าหนึ่งองค์ประกอบในอนุกรมชุดเดียวกัน ในกรณีที่ข้อมูลมีองค์ประกอบฤดูกาล การใช้ข้อมูลขนาดใหญ่มีความจำเป็นมาก เพราะถ้าข้อมูลมีจำนวนน้อยเกินไป อาจจะทำให้ไม่เห็นอิทธิพลหรือรูปแบบของฤดูกาล นอกจากนี้วิธีการบอกซ์ – เจนกินส์ จะใช้กับข้อมูลอนุกรมเวลา ที่อยู่ในสภาพนิ่งหรือคงที่ (stationary data series) ซึ่งหมายถึง คงที่ในค่าเฉลี่ย ในค่าความแปรปรวน และในความแปรปรวนร่วม (หรือสหสัมพันธ์คงที่) โดยคงที่ก็คือไม่แปรผันตามเวลา เมื่อพบว่าข้อมูลอนุกรมเวลาไม่คงที่ จำเป็นจะต้องแปลงข้อมูลให้คงที่ ก่อนที่จะดำเนินการพิจารณาสร้างตัวแบบต่อไป

การแปลงข้อมูลในกรณีที่อนุกรมเวลาไม่คงที่ในค่าเฉลี่ย จะใช้วิธีการทำผลต่าง โดยการนำข้อมูลมาลบกันได้เป็นข้อมูลชุดใหม่ ส่วนในกรณีที่อนุกรมเวลาไม่คงที่ในความแปรปรวน วิธีที่ใช้กันมากคือ ใส่ ln ในอนุกรม Y_t ได้เป็นข้อมูลอนุกรมใหม่ $X_t = \ln Y_t$ ซึ่ง $Y_t > 0$ วิธีนี้มักจะใช้เมื่อความแปรปรวนแปรผันตามระดับค่าเฉลี่ย ถ้าอนุกรมเวลาไม่คงที่ทั้งในค่าเฉลี่ย และความแปรปรวน ในกรณีนี้โดยทั่วไปจะแปลงข้อมูลในเรื่องความแปรปรวนก่อน แล้วจึงทำผลต่างในข้อมูลที่ถูกแปลงแล้ว ทั้งนี้เพราะถ้าทำผลต่างก่อน อาจแปลงแก้ความแปรปรวนไม่ได้ เช่น ถ้าผลต่างมีค่าเป็นลบ และถ้าใช้วิธีหาค่า ln จะหาค่า ln ไม่ได้ ส่วนในกรณีที่อนุกรมเวลามีองค์ประกอบของฤดูกาล โดยมีค่าเวลาของฤดูกาลหรือแปรผันตามเวลา จะแปลงข้อมูลโดยใช้วิธีการทำผลต่าง

ในการวิจัยครั้งนี้ ข้อมูลอนุกรมเวลามีองค์ประกอบฤดูกาลด้วย ดังนั้นตัวแบบที่เหมาะสมสำหรับใช้ในการพยากรณ์ของวิธีการบอกซ์ – เจนกินส์ คือตัวแบบ ARIMA เมื่อมีองค์ประกอบของฤดูกาล หรือตัวแบบ ARIMA$(p,d,q)(P,D,Q)_S$ มีรูปแบบดังนี้

$$\phi_p(B)\,\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \delta + \theta_q(B)\,\Theta_Q(B^S)\,a_t$$

โดยที่ $\Phi_P(B^S) = (1 - \Phi_S B^S - \Phi_{2S} B^{2S} - \ldots - \Phi_{PS} B^{PS})$

$\Theta_Q(B^S) = (1 - \Theta_S B^S - \Theta_{2S} B^{2S} - \ldots - \Theta_{PS} B^{PS})$

ขั้นตอนการสร้างตัวแบบ ARIMA $(p,d,q)(P,D,Q)_S$ มีดังนี้

1. กำหนดตัวแบบทดลอง ซึ่งข้อมูลอนุกรมเวลาจะต้องอยู่ในสภาพนิ่งหรือคงที่ ก่อนที่จะพิจารณาสร้างตัวแบบ โดยการพิจารณาโครงสร้างแปรผันของค่าตัวอย่าง SACF และ SPACF (ในองค์ประกอบที่เป็นฤดูกาลจะพิจารณาที่แล็ก ฤดูกาล S , 2S , . . .) ดูว่ามีลักษณะรูปแบบใกล้เคียงกับรูปแบบใด ในโครงสร้างแปรผันของค่าทางทฤษฎี ACF และ PACF มากที่สุด เพื่อจะได้เลือกกระบวนการและอันดับ เป็นตัวแบบ ARIMA ทดลอง

2. เมื่อเลือกตัวแบบ ARIMA ทดลองได้แล้ว ขั้นตอนต่อไปคือ การประมาณค่าของพารามิเตอร์ที่ปรากฏในตัวแบบ

3. วินิจฉัยความเหมาะสมเพียงพอ ของตัวแบบ ARIMA ทดลอง ด้วยการตรวจสอบคุณสมบัติ ของค่าเศษเหลือตกค้าง e_t ซึ่งเป็นค่าประมาณของค่าผิดพลาดสุ่ม a_t ว่ามีคุณสมบัติสอดคล้องกับคุณสมบัติของ a_t หรือไม่ นั่นคือค่าผิดพลาดมีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่ ไม่มีอัตสหสัมพันธ์ และมีการแจกแจงแบบปกติ หรือไม่ ถ้าพบว่าตัวแบบยังไม่เหมาะสม จะกลับไปขั้นที่ 1 พิจารณาปรับปรุงแก้ไขตัวแบบใหม่ และประมาณค่าพารามิเตอร์ในขั้นที่ 2 และวินิจฉัยตัวแบบในขั้นที่ 3 จะกระทำซ้ำ ๆ เช่นนี้ จนกว่าจะได้ตัวแบบที่เหมาะสมเพียงพอในเชิงสถิติ ที่จะใช้พยากรณ์ค่าต่อไป

3.1.2 วิธีการปรับให้เรียบแบบเลขชี้กำลัง

วิธีการพยากรณ์อนุกรมเวลา โดยวิธีการปรับให้เรียบแบบเลขชี้กำลังนี้ มีหลักการว่าจะให้ความสำคัญ หรือน้ำหนักกับข้อมูลแต่ละตัวไม่เท่ากัน โดยจะให้น้ำหนักกับข้อมูลปัจจุบันมากที่สุด และน้ำหนักจะลดลงเรื่อย ๆ สำหรับข้อมูลในอดีตที่อยู่ห่างไกลออกไปตามลำดับ ซึ่งเมื่อเขียนกราฟแสดงการลดลงของน้ำหนักจะมีรูปแบบเลขชี้กำลัง วิธีการปรับให้เรียบแบบเลขชี้กำลังมีหลายวิธี ซึ่งแต่ละวิธีจะเหมาะสมกับข้อมูลแต่ละประเภท

ในการวิจัยครั้งนี้ ข้อมูลอนุกรมเวลามีองค์ประกอบทั้งแนวโน้ม และฤดูกาล ดังนั้นวิธีการพยากรณ์ของวินเตอร์จะเหมาะสมกว่าวิธีการอื่น ๆ เนื่องจากวิธีการของวินเตอร์เหมาะสำหรับใช้ในการพยากรณ์ข้อมูลที่มีแนวโน้มและมีฤดูกาล แต่วิธีการปรับให้เรียบแบบเลขชี้กำลังอย่างง่ายเหมาะกับข้อมูลที่ไม่มีแนวโน้มและไม่มีฤดูกาล มีเฉพาะความไม่แน่นอนเพียงอย่างเดียว ส่วนวิธีการพารามิเตอร์สองตัวของโฮลท์เหมาะกับข้อมูลที่มีแนวโน้มเพียงอย่างเดียว ไม่มีฤดูกาล

3.1.3 วิธีการแยกองค์ประกอบ

วิธีการแยกองค์ประกอบเป็นการวิเคราะห์เพื่อแยกองค์ประกอบของข้อมูลอนุกรมเวลา เพื่อจะพิจารณาว่า องค์ประกอบใดที่มีอิทธิพลต่อการเปลี่ยนแปลงของข้อมูล โดยในการวิจัยครั้งนี้ข้อมูลอนุกรมเวลามีองค์ประกอบของแนวโน้ม ที่มีลักษณะเพิ่มขึ้นเป็นเส้นตรง และการผันแปรของฤดูกาลหรือความแปรปรวนมีทั้งคงที่และไม่คงที่ ดังนั้นรูปแบบของข้อมูลอนุกรมเวลาจะมี 2 รูปแบบ คือ

1. รูปแบบผลบวก (Additive Model) รูปแบบนี้เหมาะสำหรับข้อมูลอนุกรมเวลา ที่มีการแกว่ง หรือความแปรปรวน หรือการผันแปรของฤดูกาล ไม่แปรผันตามระดับค่าเฉลี่ย กล่าวคือความแปรปรวนจะคงที่ ขณะที่อนุกรมมีแนวโน้มเพิ่มขึ้นหรือมีระดับค่าเฉลี่ยสูงขึ้น ซึ่งมีรูปแบบดังนี้

$$Y_t = \beta_0 + \beta_1 t + \beta_2 I_{1t} + \ldots + \beta_{12} I_{11t} + \varepsilon_t$$

2. รูปแบบผลคูณ (Multiplicative Model) รูปแบบนี้เหมาะสำหรับข้อมูลอนุกรมเวลา ที่มีความแปรปรวน หรือการผันแปรของฤดูกาล แปรผันตามระดับค่าเฉลี่ย กล่าวคือความแปรปรวนจะมากขึ้น ขณะที่อนุกรมมีแนวโน้มเพิ่มขึ้นหรือมีระดับค่าเฉลี่ยสูงขึ้น รูปแบบผลคูณนี้จะใช้วิธีการแปลงค่าของตัวแปรตามมาช่วยแก้ปัญหา โดยการใส่ ln ในอนุกรม Y_t ซึ่งมีรูปแบบดังนี้

$$\ln Y_t = \beta_0 + \beta_1 t + \beta_2 I_{1t} + \ldots + \beta_{12} I_{11t} + \varepsilon_t$$

โดยที่ t = ตัวแปรเวลา, สำหรับ $t = 1, 2, \ldots, n$

I_{it} = ตัวแปรบ่งชี้, สำหรับ $I_{it} = 1$ เมื่อข้อมูล Y_t อยู่ในช่วงเดือนที่ i

และ $I_{it} = 0$ เมื่อข้อมูล Y_t ไม่อยู่ในช่วงเดือนที่ i

3.1.4 วิธีการวิเคราะห์การถดถอย

การวิเคราะห์การถดถอย เป็นการพยากรณ์ค่าของตัวแปรตาม Y_t ในอนาคต โดยใช้รูปแบบความสัมพันธ์ของตัวแปรตาม Y_t ที่ต้องการพยากรณ์กับตัวแปรอิสระกลุ่มหนึ่ง $X_{1t}, X_{2t}, \ldots, X_{kt}$ ($k \geq 1$) ที่ทราบค่า ซึ่งในการวิจัยครั้งนี้เป็นการหาตัวแบบการถดถอย ในกรณีวิเคราะห์รูปแบบความสัมพันธ์เชิงเส้นของตัวแปรตาม Y_t บนตัวแปรอิสระกลุ่มหนึ่ง X_{1t}, X_{2t}, X_{3t} เราเรียกรูปแบบความสัมพันธ์นั้นว่า ตัวแบบการถดถอยเชิงเส้น ซึ่งมีรูปแบบดังนี้

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t \,, \qquad \text{สำหรับ } t = 1, 2, \ldots, n$$

โดยที่ Y_t = ปริมาณความต้องการใช้ไฟฟ้าแต่ละประเภทของกลุ่มผู้ใช้ไฟฟ้าในแต่ละภาค (ล้านกิโลวัตต์ – ชั่วโมง)

X_{1t} = จำนวนประชากรในแต่ละภาค (ล้านคน)

X_{2t} = รายได้เฉลี่ยต่อคนของประชากรในแต่ละภาค (บาท)

X_{3t} = ผลิตภัณฑ์มวลรวมแต่ละประเภทของสาขาการผลิตในแต่ละภาค (ล้านบาท)

ε_t = ค่าผิดพลาดสุ่ม

ขั้นตอนการสร้างตัวแบบการถดถอยเชิงเส้น มีดังนี้

1. กำหนดรูปแบบความสัมพันธ์เชิงเส้นระหว่างตัวแปรตามและตัวแปรอิสระ โดยอาศัยข้อมูลที่มีอยู่ของตัวแปรตามและของตัวแปรอิสระ ซึ่งสามารถพิจารณาโดยใช้กราฟระหว่างตัวแปรตามกับตัวแปรอิสระทีละตัว เพื่อพิจารณากำหนดรูปแบบความสัมพันธ์เป็นคู่ ๆ ระหว่างตัวแปรตามและตัวแปรอิสระ เพื่อทดลองหาตัวแบบการถดถอยเชิงเส้นที่เหมาะสม

2. เมื่อกำหนดตัวแบบการถดถอยเชิงเส้นเป็นตัวแบบทดลองได้แล้ว ขั้นตอนต่อไปคือ การประมาณค่าพารามิเตอร์ในตัวแบบ

3. วินิจฉัยความเหมาะสมหรือความเพียงพอของตัวแบบการถดถอยเชิงเส้น ด้วยการตรวจสอบคุณสมบัติของค่าเศษเหลือตกค้าง e_t ซึ่งเป็นค่าประมาณของค่าผิดพลาดสุ่ม ε_t ว่ามีคุณสมบัติสอดคล้องภายใต้คุณสมบัติพื้นฐานของค่าผิดพลาดสุ่ม ε_t หรือไม่ นั่นคือมีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่ ไม่มีอัตสหสัมพันธ์ และมีการแจกแจงแบบปกติ หรือไม่ ถ้าตัวแบบที่กำลังพิจารณายังไม่เหมาะสม จะกลับไปทำในขั้นที่ 1 ถึง 3 ซ้ำ ๆ จนกว่าจะได้ตัวแบบที่เหมาะสมเพียงพอในเชิงสถิติที่จะใช้พยากรณ์ค่าต่อไป

4. ถ้าได้สร้างตัวแบบทดลองขึ้นมาหลายตัวแบบ และแต่ละตัวแบบได้ผ่านการวินิจฉัยแล้ว จะทำการเปรียบเทียบคัดเลือกตัวแบบที่จะใช้พยากรณ์ โดยวิธีการเปรียบเทียบค่าความถูกต้องของค่าพยากรณ์ ซึ่งจะพิจารณาจากค่า MAPE ว่าตัวแบบใดที่ให้ค่า MAPE ต่ำสุด จะเลือกตัวแบบนั้น

แผนภาพที่ 3.1 แผนภูมิแสดงโครงสร้างของระบบงานพยากรณ์