บทที่ 2

โรคที่ต้องเฝ้าระวังทางระบาดวิทยา และทฤษฎีสถิติที่เกี่ยวข้องกับงานวิจัย

ในการวิจัยครั้งนี้ได้ทบทวนเอกสาร และทฤษฎีสถิติที่เกี่ยวข้องกับการเฝ้าระวังโรคทางระบาดวิทยา โดยจะนำเสนอตามลำดับดังนี้

1. โรคที่ต้องเฝ้าระวังทางระบาดวิทยา
2. ทฤษฎีสถิติที่เกี่ยวข้องกับงานวิจัย

2.1 การเฝ้าระวังโรคทางระบาดวิทยา (กองระบาดวิทยา, 2535)

2.1.1 ประวัติความเป็นมาของการเฝ้าระวังโรคทางระบาดวิทยา

การเฝ้าระวังทางระบาดวิทยา เป็นกิจกรรมที่มีผลสืบเนื่องมาจากการพัฒนาด้านการควบคุมป้องกัน ปัญหาความเจ็บป่วยในคนตั้งแต่อดีตมาตามลำดับ โดยเริ่มตั้งแต่ในระยะแรกของการควบคุมปัญหา ที่เกิดกรณีมีการเจ็บป่วยล้มตายกันเป็นจำนวนมาก ๆ เนื่องจากโรคติดต่อร้ายแรงต่าง ๆ ซึ่งในอดีตนั้น ไม่ทราบสาเหตุ แต่เพื่อป้องกันไม่ให้คนที่ยังดีอยู่ป่วยตายไปด้วย ได้มีการใช้มาตรการ(Quarantine) โดยควบคุมตัวผู้ป่วยไว้อีกที่หนึ่งต่างหากจนกว่าโรคจะหายระบาด ครั้นต่อมาโรคติดต่อเหล่านั้นลดความรุนแรงลง และเริ่มรู้จักถึงสาเหตุของโรคมากขึ้นตามลำดับ ทำให้เข้าใจว่าไม่จำเป็นต้องเอาผู้ป่วยไปกักกันแยกจากสังคมของคนปกติ จึงมีการเปลี่ยนมาใช้มาตรการเพียงแยกกักผู้ป่วยไว้(Isolation) และเมื่อมีวิธีการด้านสาธารณสุข ได้มีการพัฒนาด้านเวชศาสตร์ป้องกันมากขึ้น มีการยอมรับว่าควรจะมีมาตรการป้องกันก่อนที่จะเกิดปัญหา ทำให้เห็นความจำเป็นที่จะต้องมีการติดตามรวบรวมข้อมูลต่าง ๆ ที่เกี่ยวกับปัญหานั้น ๆ ไว้อย่างสม่ำเสมอ ตั้งแต่เหตุการณ์ยังปกติอยู่ ในการดำเนินการดังกล่าวเรียกว่า การเฝ้าระวัง (Surveillance) ซึ่งมีอยู่ 3 ลักษณะด้วยกัน

(1) Individual Surveillance
(2) Disease Surveillance
(3) Epidemiological Surveillance

Individual or Personal Surveillance เป็นการติดตามเฝ้าระวังผู้สัมผัสโรค เพื่อให้คำแนะนำช่วยเหลือ รวมทั้งจะได้รู้ทันทีเมื่อการแสดงอาการป่วยจากโรคที่สัมผัสมา โดยไม่ต้องจำกัดการประกอบกิจกรรม หรือที่พักอาศัย ซึ่งเป็นมาตรการควบคุมป้องกันโรคอย่างหนึ่ง

Disease Surveillance เป็นการติดตามเฝ้าระวังในด้านการเกิด และการกระจายของโรค ภัยไข้เจ็บ หรือเหตุการณ์ที่เสี่ยงต่อการถ่ายทอดโรคเพิ่มสูงขึ้น ในพื้นที่ที่จำกัด

Epidemiological Surveillance เป็นการติดตามเฝ้าระวังในรายละเอียดของบุคคล สถาน ที่ และเวลา ที่เกี่ยวกับโรคภัยไข้เจ็บ และเหตุการณ์ต่าง ๆ ที่เกี่ยวกับสุขภาพของประชาชน หรือแม้แต่การเพิ่มปัญหาประชากร เพื่อให้ได้ข้อมูลในการศึกษาทางระบาดวิทยาของโรคหรือเหตุการณ์เหล่านั้น ในภาษาไทยจึงใช้ว่า "การเฝ้าระวังทางระบาดวิทยา" ซึ่งเป็นกิจกรรมที่กว้างกว่าเรื่อง "การเฝ้าระวังโรค"

2.1.2 ความหมายของการเฝ้าระวังโรคทางระบาดวิทยา

การเฝ้าระวังโรค(Surveillance) มีรากศัพท์มาจากคำว่า Vigilare ในภาษาลาติน ที่แปลว่าการเฝ้าระวังจับตาดูอย่างต่อเนื่อง สำหรับองค์การอนามัยโรคได้กล่าวไว้ว่า

"Surveillance is the continuous scrutiny of the factors that determine the occurrence and distribution of disease and other conditions of ill health. Surveillance is essential for effective control and prevention and includes the collection, analysis interpretation and distribution of relevant data for action" (WHO, 1978)

การเฝ้าระวังทางระบาดวิทยา คือ กระบวนการติดตาม สังเกต และพินิจพิจารณาลักษณะการเกิด และการกระจายของโรคภัยไข้เจ็บ ต่าง ๆ อย่างมีระบบ ซึ่งจะรวมทั้งปัจจัยต่าง ๆ ที่มีอิทธิพลต่อการเปลี่ยนแปลงต่อการเกิด และการกระจายด้วยการเฝ้าระวังทางระบาดวิทยา ประกอบด้วยขั้นตอนการรวบรวม เรียบเรียง วิเคราะห์ แปลผล และการกระจายข้อมูล เพื่อนำไปสู่การควบคุมป้องกันที่มีประสิทธิภาพ

2.1.3 วัตถุประสงค์ของการเฝ้าระวังโรคทางระบาดวิทยา

การเฝ้าระวังโรคมีวัตถุประสงค์ เพื่อ

(1) ทราบถึงการเกิดโรคภัยไข้เจ็บอย่างทันท่วงที
(2) ทราบถึงแนวโน้มการเปลี่ยนแปลงของโรค ภัย ไข้ เจ็บ
(3) ทราบถึงประชากรที่เสี่ยงต่อโรค
(4) ทราบถึงวิธีการควบคุม และป้องกันโรคภัยไข้เจ็บ

การเฝ้าระวังทางระบาดวิทยา เป็นกิจกรรมที่สำคัญทางระบาดวิทยา อีกกิจกรรมหนึ่งที่จะทำให้ทราบถึงระดับ ของการเกิดโรคในชุมชน ที่เรียกว่า โรคประจำถิ่น(Endermic) พร้อมทั้งทราบลักษณะปกติของโรคนั้น ๆ (Disease pattern) ด้วย จากการเฝ้าระวังอย่างมีประสิทธิภาพ เช่นนี้ หากมีลักษณะผิดปกติเกิดขึ้น ก็สามารถค้นพบการระบาด(Epidermic detection) ได้ทันท่วงที ไม่ระบาดแพร่กระจายออกไปวงกว้างจนไม่สามารถสอบสวนค้นหาสาเหตุการระบาดได้

2.1.4 ขั้นตอนการดำเนินงานเฝ้าระวังโรคทางระบาดวิทยา

(1) การรวบรวม(Collection) เป็นการรวบรวมข้อมูลตามองค์ประกอบต่าง ๆ โดยมีกิจกรรมย่อยดังนี้ คือ

- สังเกต(observe) พิจารณาว่าข้อมูลลักษณะผิดไปจากเดิม หรือจากข้อมูลอื่น ๆ อย่างไร
- ซักถาม(*inquire*) รวบรวมรายละเอียดของข้อมูลเพิ่มเติมให้ชัดเจนยิ่งขึ้น
- การตรวจสอบ(verify) ยืนยันให้เกิดความมั่นใจในข้อมูลที่มีอยู่ว่าถูกต้องหรือไม่เพียงใด
- บันทึก(record) จดรวบรวมรายละเอียดต่าง ๆ ไว้เป็นหลักฐาน เพื่อพิจารณาเปรียบเทียบ หรือทบทวนได้ในวันหน้า
- รายงาน(report) ส่งต่อข้อมูลให้ศูนย์รวบรวมข้อมูล

(2) การเรียบเรียง(Consolidation) นำข้อมูลที่รวบรวมไว้มาเรียบเรียงจัดหมวดหมู่เพื่อให้เห็นลักษณะของตัวแปรต่าง ๆ ได้ชัดเจนยิ่งรวมทั้งการนำข้อมูลเหล่านั้นมาประมวลเข้าด้วยกันแล้วนำเสนอทางสถิติ

(3) *การวิเคราะห์(analysis)* เป็นการนำข้อมูลมาประมวลเข้าด้วยกันตามประเด็นต่าง ๆ ของตัวแปรที่มีอยู่โดยใช้วิธีทางสถิติ เพื่อแสดงลักษณะของประชากรเจ้าของข้อมูล

(4) การแปลผล(interpretation) เป็นการนำผลการวิเคราะห์ต่าง ๆ ที่เกี่ยวข้องมาพิจารณาอย่างมีเหตุผล แล้วให้ข้อมูลสรุปหรือการวินิจฉัยที่ถูกต้อง เหมาะสมทั้งในแง่ของความเป็นไปได้ตามหลักระบาดวิทยาและชีวสถิติ

(5) การกระจายข่าวสาร(dissemination) เป็นการกระจายข้อมูล และผลการวิเคราะห์ที่แปลความหมายแล้วไปให้กับผู้ที่เกี่ยวข้องเพื่อนำไปใช้ประโยชน์ต่อไป

2.1.5 องค์ประกอบของข้อมูลการเฝ้าระวังทางระบาดวิทยา

มีทั้งสิ้น 10 องค์ประกอบ ดังนี้

(1) รายงานการป่วย เป็นรายงานจากสถานบริการสาธารณสุข ที่ให้บริการแก่ผู้ป่วย ซึ่งอาจมีจุดอ่อน ที่ผู้ให้การวินิจฉัย มีทั้งที่เป็นแพทย์และเจ้าหน้าที่สาธารณสุขอื่น ๆ ซึ่งมีความแตกต่างกันในเรื่องความรู้ และความสามารถ รวมทั้งการใช้ห้องชันสูตรช่วยยืนยันการวินิจฉัย ทำให้ความแม่นยำแตกต่างกันไป อีกทั้งรายงานการป่วยที่ได้รับรายงานอาจไม่ครบถ้วน เพราะผู้ป่วยส่วนหนึ่งไม่ได้ไปรับบริการที่สถานบริการสาธารณสุขของรัฐ โดยอาจจะไปหาแพทย์ตามคลินิคเอกชน ซื้อยากินเองจากร้านขายยา หรือไปรับการรักษาที่สถานบริการอื่น ๆ ก็ได้ ข้อมูลที่ได้รับ จึงเป็นเพียงส่วนหนึ่งของการเจ็บป่วยเท่านั้น แต่อย่างไรก็ตามข้อมูลนี้มีการเก็บอยู่เป็นประจำอยู่แล้ว การเฝ้าระวังทางระบาดวิทยา จากองค์ประกอบของข้อมูลส่วนนี้ สามารถใช้แสดงแนวโน้มของโรคที่เป็นปัญหาได้พอควร

(2) รายงานการตาย เป็นเครื่องบ่งชี้บอกถึงความรุนแรงของการเกิดโรคภัยไข้เจ็บ การบันทึกข้อมูล และการรายงานการตายใช้ใบมรณบัตร ซึ่งตามกฎหมายบังคับให้แจ้งการตายภายใน 24 ชั่วโมงต่อกำนัน หรือ นายทะเบียนท้องถิ่น อย่างไรก็ตาม การวินิจฉัยสาเหตุการตาย ยังมีปัญหาเนื่องจากมาตรฐานการวินิจฉัยไม่เหมือนกัน และผู้ให้การวินิจฉัยส่วนใหญ่ไม่มีความรู้ทางด้านการแพทย์ การเรียบเรียงและวิเคราะห์ข้อมูลเกี่ยวกับการตาย จึงควรจะพิจารณาแยกข้อมูลจากผู้รายงานแต่ละแห่งออกจากกัน เพื่อการประเมินความถูกต้อง และความน่าเชื่อถือของข้อมูลแล้วจึงทำการวิเคราะห์รายละเอียดต่อไป

(3) รายงานการชันสูตรโรค จะช่วยบ่งชี้ถึงสาเหตุโรคติดเชื้อ และปัจจัยเสี่ยง หรือความรุนแรงของโรคไร้เชื้อ การตรวจ Serotypes ของเชื้อที่มีมาในท้องที่นั้น ในช่วงเวลาหนึ่ง ความต้านทานต่อยาปฏิชีวนะของเชื้อต่าง ๆ การค้นหาผู้เป็นพาหะของโรค การตรวจพยาธิวิทยาทางชีวเคมี ทำให้ทราบถึงการเปลี่ยนแปลงของสารบางอย่างในร่างกาย จะมีความน่าเชื่อถือสูง สามารถใช้เป็นแหล่งข่าวในการติดตามสถานการณ์โรคได้อย่างดี เพียงแต่กระบวนการต่าง ๆ ต้องใช้เวลาอาจไม่ทันต่อการติดตามผู้ป่วย

(4) รายงานการสอบสวนผู้ป่วยเฉพาะราย ในกรณีที่เป็นโรคติดต่อร้ายแรง หรือโรคที่กระจายออกไปได้อย่างรวดเร็ว จะต้องสอบสวนผู้ป่วยเฉพาะราย ค้นหาสาเหตุ และแหล่งโรคเพื่อการควบคุมอย่างฉับพลัน และยังเป็นต่อการศึกษาทางระบาดวิทยา

(5) รายงานการระบาด เป็นองค์ประกอบของการเฝ้าระวังที่จะช่วยบ่งชี้ปัญหาการเกิดโรคซึ่งอาจไม่มีผู้ใดรู้สึก ผิดสังเกตมาก่อน แต่ส่วนใหญ่มักมีการรายงานน้อย แหล่งการรายงานระบาดอาจจะเป็นหน่วยงานทางสาธารณสุขในท้องถิ่น ได้แก่สถานีอนามัย โรงพยาบาล หน่วย

ชันสูตรสาธารณสุข หรืออาสาสมัครสาธารณสุข(อสม./ผสส.) ตลอดจนหน่วยงานอื่น เช่นโรงเรียน โรงงาน บ้านกำนัน บ้านผู้ใหญ่บ้าน หน่วยแพทย์ และหน่วยสาธารณสุขเคลื่อนที่ ข่าวหนังสือพิมพ์ วิทยุ และโทรทัศน์

(6) รายงานการสอบสวนการระบาดในท้องที่ ทำให้ทราบชนิดของโรคและภัยที่มีการระบาด ขอบเขตความรุนแรงของการระบาดวิทยา สาเหตุและปัจจัยของการระบาด ตลอดจนวิธีการถ่ายทอดโรค ทั้งนี้ จะต้องได้รับความร่วมมือประสานงานกันหลายฝ่าย

(7) รายงานการสำรวจทางระบาดวิทยา ให้ทราบข้อมูลพื้นฐานของปัจจัยที่เกี่ยวข้องกับการเกิดโรคที่แท้จริงในชุมชน เพื่อใช้ประกอบการติดตาม ศึกษาลักษณะของการเกิดโรค เช่น การสุ่มตัวอย่าง เจาะเลือดประชากรเป้าหมาย เพื่อตรวจภูมิคุ้มกันโรคบางชนิด การสำรวจทะเบียนโรค(Disease registers surveys) การสำรวจทะเบียนผู้ป่วยในของโรงพยาบาล(Hospital admission Surveys) เป็นต้น

(8) รายงานการศึกษาในสัตว์ และแมลงนำโรค เพื่อทราบสภาวะการเป็นรังโรคและสามารถในการนำโรคของสัตว์ และแมลงเหล่านั้น เช่นการตรวจชันสูตรโรคพิษสุนัขบ้า ในสุนัขเถื่อน หรือสัตว์ป่า หรือการสำรวจเรื่องกาฬโรค จากหมัดที่ได้จากหนูที่ดักมาจากที่ต่าง ๆ การตรวจหาเชื้อไวรัสไข้เลือดออกในยุงลาย เป็นต้น

(9) รายงานการใช้วัคซีน ซีรุ่ม และยาใช้ในท้องถิ่น เพื่อที่จะติดตามการเปลี่ยนแปลงลักษณะการเกิดโรคให้ทันต่อเหตุการณ์ ซึ่งอาจช่วยชี้แนวโน้มของสถานการณ์ได้ด้วย

(10) ข้อมูลเกี่ยวกับประชากร และสิ่งแวดล้อม ข้อมูลที่ได้จากการสำรวจสำมะโนประชากร จะช่วยในการคิดอัตราต่าง ๆ นอกจากนั้นยังมีความรู้เกี่ยวกับเศรษฐานะ การทำมาหากิน อาหาร สภาวะโภชนาการ การสุขาภิบาล แหล่งน้ำกินน้ำใช้ ชนิดและปริมาณของสัตว์เลี้ยง อุณหภูมิ ความชื้น ปริมาณฝนตก การคมนาคม แหล่งปฏิกรณ์ปรมาณู ควรมีแผนที่ของจังหวัด อำเภอ ตำบล หมู่บ้านต่าง ๆ หรือทราบแหล่งที่จะไปขอได้

2.1.6 วิธีการดำเนินงานเฝ้าระวังโรคทางระบาดวิทยา

การดำเนินงานเฝ้าระวังทางระบาดวิทยา สามารถแบ่งได้เป็น 2 วิธี คือ

(1) Active Surveillance เป็นการดำเนินงานเฝ้าระวังโรค โดยผู้รวบรวมเข้าไปตามปัญหาที่ทำการเฝ้าระวังอย่างใกล้ชิดตลอดเวลา เมื่อพบก็ทำการบันทึกรวบรวมข้อมูลทันที เช่น ตัวอย่างกรณีการเฝ้าระวังปัญหาที่จะเกิดขึ้นจากการแข่งขันกีฬาแห่งชาติ เป็นต้น วิธีการดังกล่าวจัดทำให้ทราบ ลักษณะปัญหาได้อย่างรวดเร็วและควบคุมคุณภาพของข้อมูลได้ด้วยตนเอง

ซึ่งมักจะได้ผลดีในการเฝ้าระวังในระยะสั้น ๆ และมีพื้นที่ไม่กว้างนัก โดยทั่วไปมักใช้กับโรคที่หายยาก (rare disease) แต่มีความสำคัญ และต้องการความครอบถ้วนสูง

(2) Passive Surveillance เป็นวิธีการดำเนินงานการเฝ้าระวังโดยการกำหนด ผู้บริการตามสถานบริการสาธารณสุขของรัฐ ซึ่งจะมีผู้ป่วยด้วยโรคที่อยู่ในข่ายเฝ้าระวังด้วยวิธีนี้ได้ผลดีกับการติดตามปัญหาสาธารณสุขทั่วไป และกินพื้นที่กว้าง เช่น การเฝ้าระวังปัญหาสาธารณสุขของประเทศ หรือการเฝ้าระวังปัญหาสาธารณสุขของจังหวัดต่าง ๆ เป็นต้น เนื่องจากวิธีการดังกล่าวนี้ไม่มีข้อจำกัดในด้านเวลาดำเนินการ ดังนั้นจึงเหมาะในการติดตามลักษณะการเปลี่ยนแปลงของปัญหาตามช่วงฤดูกาล หรือรอบปี จุดอ่อนที่พบบ่อยของวิธีนี้ คือ ความสม่ำเสมอ ความล่าช้าของการส่งรายงาน

2.1.7 เครื่องมือในการดำเนินงานเฝ้าระวังโรคทางระบาดวิทยา

กระทรวงสาธารณสุข มีระบบการเฝ้าระวังทางระบาดวิทยา เพื่อติดตามสถานการณ์โรคที่สำคัญ และต้องเฝ้าระวัง โดยวิธีการรายงานจากสถานบริการสาธารสุข ตั้งแต่ระดับตำบล, อำเภอ และจังหวัดมายังศูนย์กลาง คือ กองระบาดวิทยา ระบบนี้มีการควบคุมทั่วประเทศ และดำเนินการรวบรวม เรียบเรียง วิเคราะห์ แปลผล และกระจายข่าวสารสู่ผู้เกี่ยวข้องอย่างสม่ำเสมอ ในระบบเฝ้าระวังโรคทางระบาดวิทยานี้ จำเป็นต้องอาศัยระบบบันทึกในลักษณะต่าง ๆ เป็นเครื่องมือแบบบันทึกเหล่านี้ ได้แก่

(1) แบบรวบรวมข้อมูลทางระบาดวิทยาของผู้ป่วยแต่ละราย

- แบบรายงานผู้ป่วยด้วยโรคที่อยู่ในข่ายงานเฝ้าระวัง บัตรรายงานผู้ป่วย (แบบรง.506) เป็นแบบรายงานที่ใช้บันทึกข้อมูลของผู้ป่วย แต่ละรายที่ป่วยด้วยโรคที่อยู่ในข่ายงานเฝ้าระวัง แบบรง.506 ยังมีข้อมูลผู้ป่วยในด้านลักษณะบุคคล สถานที่เริ่มป่วย และวันที่เริ่มป่วยและอื่น ๆ อีกด้วย หลักการสำคัญในการบันทึกแบบรายงายนี้ คือ แม้แต่สงสัยก็ให้รายงาน ไม่จำเป็นต้องรอผลชันสูตรเพื่อยืนยันการวินิจฉัย (ยกเว้นบางโรคที่กำหนดเป็นพิเศษ ทั้งนี้เพื่อความรวดเร็วในการรายงาน และตรวจจับการระบาด)

- บัตรเปลี่ยนแปลงการรายงานผู้ป่วย (แบบ รง.507) หลังจากส่ง รง. 506 ไปแล้ว อาจมีการเปลี่ยนแปลงภายหลังเกี่ยวกับผู้ป่วย จึงได้มีแบบบันทึกเพื่อแก้ไขรายงาน คือ บัตรเปลี่ยนแปลงการรายงานผู้ป่วย (แบบ รง.507) หน้าที่ที่เกี่ยวข้องเนื่องจากต้องอาศัยระบบบริหารจัดการข้อมูลที่มีประสิทธิภาพ และการประสานงานที่ดีระหว่างฝ่ายที่เกี่ยวข้องภายในสถานบริการ และระหว่างหน่วยงาน จึงจะกรอกรายงาน 507 ได้ ครบถ้วนทุกตัวแปรที่ต้องการได้ บัตรนี้มีความสำคัญมาก เพราะเป็นเครื่องมือหนึ่งที่จะช่วยให้ข้อมูลเฝ้าระวังโรคน่าเชื่อถือ และมีความ

ถูกต้องสูง อย่างไรก็ตามทั้งแบบ รง.506 และ รง.507 มีหน้าที่เป็นเพียงเครื่องมือในการรวบรวมข้อมูลซึ่งเป็นเพียงขั้นตอนหนึ่งของการเฝ้าระวังทางระบาดวิทยา

(2) แบบเรียบเรียง

เนื่องจากตัวแปรทางระบาดวิทยาที่สำคัญมีหลายตัวแปรด้วยกันแบบพิมพ์ที่ใช้เป็นเครื่องมือในการเรียบเรียงข้อมูล จึงมีหลายลักษณะด้วยกัน ดังนี้

- แบบ E.1 ใช้เรียบเรียงข้อมูล โดยแยกเป็นรายโรค ใน 1 ชุด จะมีข้อมูลลักษณะของผู้ป่วยที่ป่วยด้วยโรคเดียวกันทั้งหมดโดยละเอียด เช่น ถ้าอยากทราบว่าในแต่ละโรคมีจำนวนผู้ป่วยมากน้อยเพียงใดก็สามารถบอกได้จากช่องลำดับของผู้ป่วยโรคนั้น ๆ E.1 ซึ่งเป็นเครื่องมือเรียบเรียงข้างต้น

- แบบ E.2 ใช้เรียบเรียงข้อมูลโดยแยกจำนวนผู้ป่วยออกตามสถานเริ่มป่วย ในช่วงเวลาแต่ละเดือนในรอบปี สำหรับสถานที่เริ่มป่วยอาจแยกเป็นพื้นที่ ตำบล อำเภอ หรือจังหวัดก็ได้ ทั้งนี้ขึ้นอยู่กับระดับของข่ายงานเฝ้าระวังระวังทางระบาดวิทยาที่วิเคราะห์ข้อมูลในขั้นต่อไป ผลการเรียบเรียงข้อมูลตามแบบ E.2 นี้ จะทำให้เห็นลักษณะการกระจายของปัญหาไปตามพื้นที่ต่าง ๆ ซึ่งหากทำการวิเคราะห์ต่อ ไปอีกเพียงเล็กน้อยก็สามารถระบุพื้นที่ที่เสี่ยงต่อปัญหาในแต่ละช่วงเดือนได้

- แบบ E.3 ใช้เรียบเรียงข้อมูล โดยแยกจำนวนผู้ป่วยออกจากกลุ่มอายุและเพศในช่วงเวลาแต่ละเดือนในรอบปี เช่นเดียวกับแบบ E.2 ดังนั้น จึงสามารถบอกลักษณะการกระจายของปัญหาไปตามกลุ่มคนกลุ่มต่าง ๆ และหากทำการวิเคราะห์ต่อไป ก็สามารถระบุกลุ่มประชากรที่เสี่ยงต่อปัญหานั้น ๆ ได้เช่นกัน

- แบบบันทึกผู้ป่วยประจำวัน(Daily record) ใช้เรียบเรียงข้อมูล โดยแยกจำนวนผู้ป่วยออกตามสถานที่เริ่มป่วยแต่ละวันใน 1 รอบเดือน ทำให้สามารถบอกความผิดปกติของจำนวนผู้ป่วยแต่ละพื้นที่ได้ทันทีที่เริ่มมีการเพิ่มขึ้นของจำนวนผู้ป่วย และเมื่อเทียบกับจำนวนในช่วงเวลาเดียวกันของปีก่อน ๆ

- แบบ E.4 ใช้เรียบเรียงข้อมูล โดยแยกจำนวนผู้ป่วยออกตามโรค ที่เป็นปัญหาอยู่ในข่ายงานเฝ้าระวังตามพื้นที่ต่าง ๆ แบบ E.4 นี้มีลักษณะพิเศษ กล่าวคือ การนับจำนวนผู้ป่วยนั้น นับจากวันรักษาในรอบ 1 สัปดาห์ ที่ผ่านมา ไม่ใช่นับวันเริ่มป่วย ผลที่ได้เป็นเพียงลักษณะความผิดปกติจากจำนวนผู้ป่วยที่มาทำการรักษาเพิ่มมากขึ้น ซึ่งอาจเกิดปัญหาการระบาดหรือไม่ก็ได้ต้องตรวจสอบจาก Daily record ต่อไป จึงจะบอกให้แน่ชัด

2.1.8 ข่ายงานเฝ้าระวังทางระบาดวิทยา

เพื่อให้สามารถติดตามปัญหาสาธารณสุขครอบคลุมได้ทั้งประเทศจึงได้จัดระบบการเฝ้าระวังเป็นลักษณะเครือข่ายในระดับต่าง ๆ โดยมีองค์กรรับผิดชอบ และหน่วยงานที่เกี่ยวข้องในแต่ละระดับดังนี้

(1) ระดับประเทศ
- สำนักงานปลัดกระทรวงสาธารณสุข
- กองระบาดวิทยา
- ศูนย์ระบาดวิทยาภาคฯ
  - ภาคเหนือ จังหวัดลำปาง
  - ภาคตะวันออกเฉียงเหนือ จังหวัดนครราชสีมา
  - ภาคกลาง จังหวัดราชบุรี
  - ภาคใต้ จังหวัดสงขลา

(2) ระดับจังหวัด
- สำนักงานสาธารณสุขจังหวัด
  - งานระบาดวิทยา (เจ้าหน้าที่ระบาดวิทยาประจำจังหวัด)
- โรงพยาบาลศูนย์/โรงพยาบาลทั่วไป
  - ฝ่ายเวชกรรมสังคม

(3) ระดับอำเภอ
- สำนักงานสาธารณสุขอำเภอ
  - งานสุขาภิบาลและป้องกันโรค
- โรงพยาบาลชุมชน
  - งานสุขาภิบาลและป้องกันโรค

(4) ระดับตำบล
- สถานีอนามัย
  - งานเฝ้าระวังและควบคุมโรค

2.2 ทฤษฎีสถิติที่เกี่ยวข้องในงานวิจัย

การวิเคราะห์อนุกรมเวลา (Time Series Analysis) เป็นการศึกษาข้อมูลที่ให้ความสำคัญต่อ ลำดับที่เกิดขึ้นของข้อมูลนั้น และจะเน้นถึงความสำคัญที่ขึ้นต่อกัน (Dependence) ของข้อมูล

จากลักษณะทั้ง 2 ประการนี้เองที่ทำให้การวิเคราะห์อนุกรมเวลาแตกต่างจากกรรมวิธีทางสถิติอื่น ๆ ที่มีข้อมูลสมมติเกี่ยวกับความเป็นอิสระ (Independence) ต่อกัน และการเกิดสุ่ม (Randomization) ของข้อมูล จุดมุ่งหมายที่สำคัญของการวิเคราะห์อนุกรมเวลา คือการอธิบายขบวนการที่ก่อให้เกิดอนุกรมเวลาชุดนั้น ๆ และพยากรณ์ค่าของตัวแปรในอนาคต โดยถือว่าค่าพยากรณ์ในอนาคตเป็นผลที่พิจารณาได้จากข้อมูลในอดีต

วิธีการวิเคราะห์อนุกรมเวลาที่ศึกษาครั้งนี้มีวิธีวิเคราะห์ที่ใช้ 5 วิธีคือวิธีพยากรณ์โดยใช้วิธีของบอกซ์และเจนกินส์, เทคนิคการปรับให้เรียบ, การวิเคราะห์อนุกรมเวลาแบบคลาสสิค, วิธีการหาค่าพยากรณ์ร่วมด้วยวิธีการให้เฉลี่ยถ่วงน้ำหนักด้วยวิธีการให้น้ำหนักที่เท่ากัน และวิธีค่าสัมบูรณ์ต่ำสุด ซึ่งมีวิธีการดังต่อไปนี้

2.2.1 วิธีการของบอกซ์และเจนกินส์

วิธีการพยากรณ์อนุกรมเวลาโดยทั่วไป จะมีข้อสมมติฐานข้อหนึ่งคือ อนุกรมเวลา $\{..., Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1},...\}$ ไม่มีสหสัมพันธ์ ดังเช่นวิธีการปรับให้เรียบที่ว่าไม่เป็นจริง นั่นคือ มีหลายกรณีที่อนุกรมเวลา $\{..., Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1},...\}$ มีสหสัมพันธ์ถ้าเป็นเช่นนี้ การใช้วิธีการปรับให้เรียบ หรือวิธีอื่นที่มีข้อสมมติตัวแปรอนุกรมเวลาไม่มีสหสัมพันธ์อาจจะไม่เหมาะสม ทั้งนี้เพราะวิธีการเหล่านี้ไม่ได้นำเอาสหสัมพันธ์ ที่ปรากฏไปใช้ประโยชน์ ในการสร้างตัวแบบพยากรณ์ ฉะนั้นได้มีผู้คิดค้นหาวิธีการสร้างตัวแบบพยากรณ์สำหรับอนุกรมเวลาที่ได้นำเอาสหสัมพันธ์ที่ปรากฏไปวิเคราะห์ใช้ประโยชน์ และโดยทั่วไปวิธีเหล่านี้จะให้ผลพยากรณ์ที่ดีกว่า ในหลายวิธีเหล่านี้ วิธีหนึ่งที่เป็นที่รู้จักและใช้กันมากคือวิธีบอกซ์และเจนกินส์

โดยวิธีการบอกซ์และเจนกินส์ จะหาตัวแบบอนุกรมเวลาโดยพิจารณาสหสัมพันธ์ระหว่าง Y ที่ตำแหน่งเวลาหรือคาบเวลา t (Y_t) และ Y ที่ตำแหน่งเวลา หรือคาบเวลาต่าง ๆ ที่ผ่านมา $(Y_{t-2}, Y_{t-1}, ...)$ เมื่อได้ตัวแบบแล้วตัวแบบนี้จะแสดงความสัมพันธ์ระหว่าง Y_t กับ $Y_{t-1}, Y_{t-2}, ...$ และจะใช้ตัวแบบนี้พยากรณ์ค่า $Y_{t-1}, Y_{t-2}, ...$ ในอนาคต ตัวแบบบอกซ์และเจนกินส์โดยทั่วไป จะใช้พยากรณ์ค่าในช่วงเวลาข้างหน้าที่เป็นระยะสั้นหรือปานกลาง ทั้งนี้เพราะตัวแบบโดยทั่วไปจะให้ความสำคัญ หรือน้ำหนักมากกว่าข้อมูลอดีตที่ห่างไกลออกไปมาก ๆ เราอาจจะไม่เคยพบตัวแบบบอกซ์และเจนกินส์ ที่แสดงความสัมพันธ์ระหว่าง Y_t กับ Y ที่ห่างไกลกันมาก ๆ

ลักษณะตัวแบบบอกซ์และเจนกินส์มาจากการศึกษาวิเคราะห์กระบวนการเชิงเส้นหรือตัวกรองเชิงเส้น (linear filter):

$$Y_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots \quad (2.1)$$

นั่นคือ พิจารณาอนุกรมเวลาหรือค่าสังเกต Y_t เกิดจากผลบวกเชิงเส้นของอนุกรมเวลาค่าผิดพลาดสุ่ม a_t , a_{t-1} , a_{t-2} ซึ่งสมมติว่าแต่ละตัวมีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่ และไม่มีสหสัมพันธ์กัน หรือมีความแปรปรวนร่วม $\mathrm{Cov}(a_t, a_{t-k}) = 0$ สำหรับทุกค่า $k \neq 0$ และโดยทั่วไปสมมติด้วยว่ามีการแจกแจงแบบปกติ

ในตัวกรองเชิงเส้น หรือตัวแบบเชิงเส้น (2.1) พารามิเตอร์ μ คือค่าระดับค่าเฉลี่ยของ Y_t เมื่ออนุกรมเวลาอยู่ในสภาพคงที่และพารามิเตอร์ ψ_1 , ψ_2 ,... เป็นน้ำหนักที่ให้กับตัวแปรสุ่ม a_{t-1} ,a_{t-2},... กระบวนการหรือตัวแบบเชิงเส้น (2.1) จะไม่ให้ประโยชน์ถ้ามีพารามิเตอร์จำนวนอนันต์ เพราะฉะนั้นจะสร้างตัวแบบที่ประกอบด้วยพารามิเตอร์จำนวนจำกัด และเพียงพอที่จะอธิบายอนุกรมเวลาที่พิจารณา

2.2.1.1. **ตัวแบบภายใต้ภาวะคงที่** (Stationary Models)

2.2.1.1.1 **ตัวแบบอัตถดถอย** (Autoregressive (AR) Models)

จากตัวแบบเชิงเส้น (2.1) พัฒนาตัวแบบเฉพาะกลุ่มหนึ่งเรียกว่า " ตัวแบบอัตถดถอยอันดับ p หรือกระบวนการอัตถดถอยอันดับ p " (Autoregessive Process of order p) แทนด้วยตัวอักษร AR (p) และมีรูปแบบทั่วไปดังนี้

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + a_t$$

หรือ

$$Z_t = \phi_1 Z_{t-1} + \phi Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t \quad (2.2)$$

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + a_t$$

โดยให้

$$Z_t = Y_t - \mu, \quad Z_{t-1} = Y_{t-1} - \mu, \ldots, c = \mu(1 - \phi_1 - \phi_2 \ldots - \phi_p)$$

และ $\phi_1, \phi_2, \ldots, \mu$ เป็นพารามิเตอร์ซึ่งโดยทั่วไปไม่ทราบค่า ต้องประมาณจากข้อมูล

จากตัวแบบ AR (p) ข้างต้นอาจเขียนใหม่ในรูปแบบสั้น ๆ

$$\phi_p(B) Z_t = a_t \quad (2.3)$$

โดยที่

$$\phi_p(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right)$$

และ

$$\begin{aligned} BZ_t &= Z_{t-1} \\ B^2 Z_t &= Z_{t-2} \\ \ldots &= \ldots \\ B^p Z_t &= Z_{t-p} \end{aligned}$$

เราเรียก B ว่า "ตัวถอยหลังเวลา" (Backward - Shift Operator)

โดยทั่วไปในทางปฏิบัติ อันดับ p จะไม่สูงมากถ้า p = 1,2 เขียนกระบวนการคงที่หรือตัวแบบคงที่ AR(1) และ AR(2) ได้ดังนี้

ตัวแบบ AR (1) $Y_t = c + \phi_1 Y_{t-1} + a_t$

โดยมีเงื่อนไขของการเป็นตัวแบบคงที่ ดังนี้ $-1 < \phi_1 < 1$

ตัวแบบ AR (2) $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t$

โดยมีเงื่อนไขของการเป็นตัวแบบคงที่ ดังนี้

$$\phi_1 + \phi_2 < 1$$

$$\phi_1 - \phi_2 < 1$$

$$-1 < \phi_1 < 1$$

2.2.1.2 **ตัวแบบค่าเฉลี่ยเคลื่อนที่** (Moving Average (MA) Model)

จากตัวแบบเชิงเส้น (2.1) พัฒนาตัวแบบเฉพาะอีกกลุ่มหนึ่งเรียกว่า "ตัวแบบค่าเฉลี่ยเคลื่อนที่อันดับ "q" หรือ "กระบวนการค่าเฉลี่ยเคลื่อนที่อันดับ "q" (Moving Average Process of order q) แทนด้วยอักษรย่อ MA(q) และมีตัวแบบทั่วไปดังนี้

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \qquad (2.6)$$

ตัวแบบบอกซ์และเจนกินส์มีเงื่อนไขที่ต้องสอดคล้องอีกเงื่อนไขหนึ่งนอกเหนือจากเงื่อนไขคงที่ "ผกผันได้" (invertibility) ซึ่งพบว่าตัวแบบ AR มีคุณสมบัติผกผันได้เสมอ แต่อาจไม่คงที่ขณะ

ที่ตัวแบบ MA มีคุณสมบัติคงที่เสมอแต่อาจจะผกผันไม่ได้ เพราะฉะนั้น ต้องตรวจสอบคุณสมบัติคงที่ในตัวแบบ AR และตรวจสอบคุณสมบัติผกผันได้ในตัวแบบ MA

ทำนองเดียวกันกับตัวแบบ AR ในทางปฏิบัติจะมีอันดับ q ไม่สูงมาก เช่น 1, 2 หรือ 3 ถ้า q = 1 และ 2 จะได้ตัวแบบ MA (1) และ MA (2) ดังนี้

ตัวแบบ MA (1)

$$Y_t = \mu + a_t - \theta_1 a_{t-1}$$

โดยมีเงื่อนไขผกผันได้ ดังนี้ $-1 < \theta_1 < 1$

ตัวแบบ MA(2)

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

โดยมีเงื่อนไขผกผันได้ ดังนี้ $\theta_1 + \theta_2 < 1$

$\theta_1 - \theta_2 < 1$

$-1 < \theta_2 < 1$

2.2.1.3 **ตัวแบบอัตถดถอย-ค่าเฉลี่ยเคลื่อนที่** (Autoregressive-Moving Average (ARMA) Model)

ในบางกรณี การใช้ตัวแบบหรือกระบวนการผสมระหว่างตัวแบบ AR และMA เป็นตัวแบบที่ประหยัด แทนการใช้ตัวแบบ AR อันดับสูง ๆ ตัวแบบเดียว หรือแทนการใช้ตัวแบบ MA อันดับสูง ๆ ตัวแบบเดียวเราจะใช้สัญลักษณ์ ARMA (p,q) หมายถึงตัวแบบผสมอัตถดถอยอันดับ p และค่าเฉลี่ยที่อันดับ q และมีรูปแบบดังนี้

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

หรือ

$$\phi(B) Z_t = \theta_q(B) a_t$$

อันดับ p และ q สำหรับตัวแบบ ARMA จะไม่สูงมากในทางปฏิบัติถ้า $p = 1$ และ $q = 1$ จะได้ตัวแบบ ARMA (1,1)

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + a_t - \theta_1 a_{t-1}$$

หรือ

$$Y_t = c + \phi_1 Y_{t-1} - a_t - \theta_1 a_{t-1}$$

โดยมีเงื่อนไขของความคงที่และผกผันได้ดังนี้

$$-1 < \phi_1 < 1 \quad \text{และ} \quad -1 < \theta_1 < 1$$

2.2.1.4 ตัวแบบภายใต้ภาวะไม่คงที่ (Nonstationary Models) และตัวแบบ ARIMA

ถ้าข้อมูลอนุกรมเวลาหรือกระบวนการ Y_t ไม่อยู่ในสภาพคงที่ในค่าเฉลี่ย และ/หรือความแปรปรวน นักพยากรณ์จะต้องแปลงข้อมูลนั้นให้อยู่ในสภาพคงที่ก่อนพิจารณากำหนดตัวแบบ

การแปลงข้อมูลอนุกรมเวลาที่ไม่คงที่ในค่าเฉลี่ย จะใช้วิธีการทำผลต่าง โดยนำข้อมูลมาลบกันได้เป็นข้อมูลชุดใหม่ นอกจากนี้การทำผลต่างดังกล่าว อาจจะต้องทำมากกว่าหนึ่งครั้งจึงจะให้อนุกรมเวลามีค่าเฉลี่ยคงที่ ซึ่งโดยทั่วไปถ้าอนุกรมมีแนวโน้มมักจะทำผลต่างสองครั้งจึงจะคงที่ การทำผลต่างไม่ควรทำหลายครั้งมากเกินความจำเป็น เพราะจะมีผลทำให้ค่าพยากรณ์มีความคลาดเคลื่อนสูง

เมื่อทำผลต่างครั้งที่หนึ่ง จะได้ข้อมูลชุดใหม่เท่ากับ

$$W_t = (1 - B)Y_t = Y_t - Y_{t-1} \qquad , t = 2, 3, \ldots, n$$

และถ้าทำครั้งที่สอง จะได้ข้อมูลชุดใหม่

$$\begin{aligned} X_t &= (1 - B)W_t = (1 - B)^2 Y_t \\ &= (1 - 2B + B^2)Y_t \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \qquad , t = 3, 4, \ldots, n \end{aligned}$$

และถ้าทำครั้งที่ d (ทำ d ครั้ง) จะได้ข้อมูลชุดใหม่ ให้เป็น W_t

$$W_t = (1 - B)^d Y_t \qquad , t = d + 1, d + 2, \ldots, n$$

ในกรณีที่อนุกรมเวลาไม่คงที่ในความแปรปรวน หรือมีการเคลื่อนไหวเป็นเส้นโค้ง วิธีการแปลงข้อมูลที่ใช้กันมากคือ ใส่ ***ln*** ในอนุกรมเวลา Y_t ได้เป็นข้อมูลอนุกรมเวลาใหม่ $X_t = \ln Y_t$ ซึ่ง $Y_t > 0, t = 1, 2, \ldots, n$ วิธีนี้มักจะใช้เมื่อความแปรปรวนแปรผันตามระดับค่าเฉลี่ย เช่น ความแปรปรวนมากขึ้นขณะที่อนุกรมมีแนวโน้มเพิ่มขึ้น หรือมีระดับค่าเฉลี่ยสูงขึ้น บางกรณีการใช้

ln อาจไม่ได้ผลนักพยากรณ์ควรทดลองใช้วิธีอื่นเช่น ใช้วิธีหารากที่สองได้ $X_t = \sqrt{Y_t}$ ซึ่ง $Y_t > 0, t = 1, 2, \ldots, n$

ถ้าอนุกรมเวลาไม่คงที่ทั้งในค่าเฉลี่ย และความแปรปรวน ในกรณีนี้จะต้องแปลงข้อมูลให้คงที่ในค่าเฉลี่ย และความแปรปรวน และโดยทั่วไปจะแปลงข้อมูลในเรื่องความแปรปรวนก่อน แล้วจึงทำผลต่างในข้อมูลที่ถูกแปลงแล้ว ทั้งนี้เพราะ ถ้าผลต่างก่อนอาจแปลงแก้ความแปรปรวนไม่ได้ เช่น ถ้าผลต่างมีค่าเป็นลบ และถ้าใช้วิธีหาค่า ***ln*** หรือรากที่สอง จะหาค่า ***ln*** หรือรากที่สองไม่ได้

เมื่ออนุกรมเวลามีสภาพไม่คงที่ หรือไม่เคลื่อนไหวรอบค่าเฉลี่ยคงที่ค่าหนึ่งค่าเดียวจะต้องแปลงข้อมูลที่กล่าวไปแล้ว ฉะนั้น ถ้ามีการทำผลต่าง d ครั้ง จะเขียนตัวแบบผสมเป็น ARIMA (Autoregressive Integrated Moving Model) ด้วยอันดับ (p,d,q) มีรูปแบบทั่วไปดังนี้

ตัวแบบ ARIMA (p,d,q)

$$\phi_p(B)(1-B)^d Y_t = \delta + \theta_q(B)a_t$$

หรือ

$$\phi_p(B)W_t = \delta + \theta_q(B)a_t$$

ซึ่งให้ $W_t = (1-B)^d Y_t$ และ δ (อาจจะมีค่าเป็นศูนย์) เป็นพารามิเตอร์แสดงระดับค่าเฉลี่ยคงที่ของอนุกรม W_t และ Y_t เป็นอนุกรมที่ถูกแปลงให้มีความแปรปรวนคงที่แล้ว ถ้าอนุกรมแรกเริ่มไม่คงที่ในความแปรปรวน ตัวอย่างต่อไปนี้เป็นตัวอย่างตัวแบบ ARIMA (p,d,q)

ตัวแบบ ARIMA (1,1,1) มีรูปแบบดังนี้

$$(1-\phi_1 B)(1-B)Y_t = \delta + (1-\theta_1 B)a_t$$

หรือ

$$Y_t = \delta + (1+\phi_1)Y_{t-1} - \phi_1 Y_{t-2} + a_t - \theta_1 a_{t-1}$$

หรือ

$$W_t = \delta + \phi_1 W_{t-1} + a_t - \theta_1 a_{t-1}, \quad W_t = Y_t - Y_{t-1}$$

ตัวแบบ ARIMA (2,1,0) มีรูปแบบดังนี้

$$(1-\phi_1 B-\phi_2 B^2)(1-B)Y_t = \delta + a_t$$

หอสมุดกลาง ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หรือ

$$W_t = \delta + \phi_1 W_{t-1} + \phi_2 W_{t-2} + a_t \ , \quad W_t = Y_t - Y_{t-1}$$

ตัวแบบ ARIMA (1,2,1) มีรูปแบบดังนี้

$$(1 - \phi_1 B)(1 - B)^2 Y_t = \delta + (1 - \theta_1 B) a_t$$

หรือ

$$W_t = \delta + \phi_1 W_{t-1} + a_t - \phi_1 W_{t-1} \ , \quad W_t = (1 - B)^2 Y_t$$

2.2.1.5 ตัวแบบ ARIMA เมื่อมีองค์ประกอบฤดูกาล

องค์ประกอบที่เป็นฤดูในที่นี้ จะไม่จำกัดเฉพาะการแปรผันของข้อมูลคล้ายคลึงกันในระยะเวลาห่างกัน 1 ปี (คล้ายคลึงกันจากปีหนึ่งไปยังปีหนึ่ง) แต่จะรวมถึงความแปรผันคล้ายคลึงกันในช่วงเวลาห่างอื่น ๆ ด้วย

สำหรับอนุกรมเวลาที่มีองค์ประกอบฤดูกาลโดยมีคาบเวลาของฤดูกาล $s > 1$ ตัวแบบอนุกรมเวลาในส่วนที่เป็นฤดูกาลจะมีโครงสร้างเป็นไปได้เหมือนกับองค์ประกอบที่ไม่ใช่ฤดูกาลนั่นคือจะมีตัวแบบ ARIMA ด้วยอันดับ (P,D,Q) ซึ่ง P คืออันดับในส่วนของกระบวนการ AR ,Q คืออันดับในส่วนของกระบวนการ MA และ D คือจำนวนครั้งทำผลต่างอนุกรมเวลาห่างกัน s คาบเวลา

เมื่อนำองค์ประกอบในส่วนที่ไม่ใช่ฤดูกาล และส่วนที่เป็นฤดูกาลมาผนวกเข้าด้วยกันจะได้ตัวแบบ ARIMA ที่แสดงส่วนประกอบทั่วไป เขียนเป็นสมการได้ดังนี้ ตัวแบบหนึ่งคือตัวแบบในรูปผลคูณ ARIMA $(p,d,q)\ (P,D,Q)_s$

ตัวแบบ ARIMA $(p,d,q)\ (P,D,Q)_s$

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

โดยที่

$$\Phi_P(B^s) = (1 - \Phi_s B^s - \Phi_{2s} B^{2s} - \ldots - \Phi_{Ps} B^{Ps})$$
$$\Theta_Q(B^s) = (1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \ldots - \Theta_{Qs} B^{Qs})$$

4. ขั้นตอนการสร้างตัวแบบบอกซ์และเจนกินส์ หรือตัวแบบ ARIMA

กรรมวิธีสร้างตัวแบบบอกซ์และเจนกินส์ หรือ ตัวแบบ ARIMA จะมีขั้นตอนใหญ่ ๆ ทำนองเดียวกันกับวิธีสร้างตัวแบบการถดถอย กล่าวคือ ประกอบด้วยขั้นตอน

1) กำหนดตัวแบบทดลอง (Identification) คือ กำหนดตัวแบบที่พิจารณาว่าจะเป็นตัวแบบที่เหมาะสมหรือมีความเพียงพอในเชิงสถิติ
2) จากตัวแบบที่เลือก ประมาณค่าองค์ประกอบหรือค่าพารามิเตอร์ของตัวแบบ (Estimation)
3) วินิจฉัยตัวแบบ (Diagnostic checking) ตัวแบบที่เลือกและประมาณค่าพารามิเตอร์แล้ว อาจยังไม่มีความเหมาะสมเพียงพอในเชิงสถิติจึงควรตรวจสอบวินิจฉัยตัวแบบ ซึ่งถ้าไม่ผ่านการตรวจสอบวินิจฉัย ให้กลับไปขั้นตอนที่ 1 ใหม่ จนกว่าจะได้ตัวแบบที่เหมาะสม

4.1 การกำหนดตัวแบบ ARIMA $(p,d,q)\ (P,D,Q)_s$

การกำหนดตัวแบบ ARIMA นักพยากรณ์จะต้องพิจารณากำหนดอันดับ p,d และ q และต้องกำหนดอันดับ P,D,Q และ s ด้วยถ้าตรวจพบว่าข้อมูลอนุกรมเวลามีองค์ประกอบฤดูกาลด้วยคาบฤดูกาล s

อันดับ p และ q คือ อันดับของกระบวนการ AR และ MA ในส่วนที่ไม่ใช่ฤดูกาล และ P และ Q คือ อันดับของกระบวนการ AR และ MA ในส่วนที่เป็นฤดูกาลสำหรับ d คือจำนวนครั้งทำผลแตกต่างอนุกรมเวลาเมื่ออนุกรมเวลาในส่วนที่ไม่ใช่ฤดูกาลมีสภาพไม่คงที่ในค่าเฉลี่ย และ D คือ จำนวนครั้งทำผลแตกต่างอนุกรมเวลาเมื่ออนุกรมเวลาในส่วนที่เป็นฤดูกาลมีสภาพไม่คงที่ในค่าเฉลี่ย

การพิจารณากำหนดอันดับ (p,d,q) และ $(P,D,Q)_s$ จะพิจารณาแยกจากกันแต่ใช้หลักการพิจารณาเหมือนกันกระบวนการ AR และ MA ต่างมีรูปแบบโครงสร้างเฉพาะสำหรับอันดับ p และ q ของฟังก์ชันอัตสหสัมพันธ์ ACF(Autocorrelation Function) แทนด้วย ϕ_{kk} ซึ่ง k หมายถึงคาบเวลาห่างระหว่างอนุกรม และเรียกคาบเวลานี้ว่าแล็ก k (lag k) ฉะนั้น ρ_1 หมายถึงอัตสหสัมพันธ์ที่แล็ก 1 หรือ อัตสหสัมพันธ์ของอนุกรมเวลาที่ห่างกัน 1 หน่วย หรือ 1 คาบเวลา $(Y_t, Y_{(t+1)})$; t = 1,2,... ซึ่งวัดความสัมพันธ์เชิงเส้นระหว่างอนุกรมเวลาที่ห่างกัน 1 คาบเวลา และ ρ_2 คือ

อัตสหสัมพันธ์ของอนุกรมเวลาที่ห่างกัน 2 คาบเวลาสำหรับ (Y_t, Y_{t+2}) , t = 1,2,... สำหรับ ϕ_{kk} เป็นอัตสหสัมพันธ์ที่วัดความสัมพันธ์ระหว่างอนุกรมเวลาที่ห่างกัน k คาบเวลา (Y_t, Y_{t+k}) โดยพิจารณาจากผลกระทบจากอนุกรมเวลา $Y_t, Y_{t+2}, \ldots, Y_{t+k-1}$ เข้าด้วยกัน ค่าของ ρ_k และ 0 ต่างมีค่าอยู่ระหว่าง −1 และ 1 ตัวอย่างสูตรของฟังก์ชันเหล่านี้เช่น

กระบวนการ AR (1) :

$$Y_t = \delta + \phi Y_{t-1} + a_t$$

$$\rho_k = \phi^k \quad ,k = 0,1,\ldots$$

$$\phi_{11} = \rho_1 = 0 \quad ,k = 2,3,\ldots$$

กระบวนการ MA(1) :

$$Y_t = \mu + a_t - \theta a_{t-1}$$

$$\rho_1 = \frac{-\theta}{1+\theta^2}, \quad \rho_k = 0 \quad ,k = 2,3,\ldots$$

$$\phi_{kk} = \frac{-\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} \quad ,k = 1,2,\ldots$$

เพราะฉะนั้นการกำหนดอันดับจะประมาณค่า ρ_k และ ϕ_{kk} โดยใช้ข้อมูลอนุกรมเวลาที่นำมาวิเคราะห์ แทนค่าประมาณด้วย $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ และเรียกว่า "ฟังก์ชันอัตสหสัมพันธ์ตัวอย่าง" และ "ฟังก์ชันอัตสหสัมพันธ์ย่อยตัวอย่าง" SACF (Sample Autocorrelation Function) และ "ฟังก์ชันอัตสหสัมพันธ์ย่อยตัวอย่าง" SPACF (Sample Partial Autocorrelation Function)

ค่าประมาณ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ ซึ่งจะคำนวณค่าโดยใช้ข้อมูลอนุกรมเวลามีสูตรทั่วไปดังนี้

$$\hat{\rho}_k = \sum_{t=1}^{n-k} \frac{(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \quad ,k = 0,1,2,\ldots$$

$$\phi_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_{kk} - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}(\hat{\rho}_{k-j})}{1 - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}(\hat{\rho}_j)} \quad ,k = 2,3,\ldots.$$

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j} \quad (k = 3,4,\ldots; j = 1,2,\ldots,k-1)$$

Y_t = ข้อมูลที่เวลา t

การคำนวณค่าที่ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ แล็ก k=1,2,...,m เพื่อให้มีจำนวนค่ามากพอที่จะพิจารณาโครงสร้างการแปรผันของอัตสหสัมพันธ์ได้ง่าย จะคำนวณค่า $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ แต่ละประเภทจำนวน m = n/4 โดยประมาณจากค่าประมาณเหล่านี้จะเปรียบเทียบลักษณะแปรผันกับโครงสร้างแปรผันของ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ ทางทฤษฎีและเลือกตัวแบบ ARIMA ที่โครงสร้างของ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ เข้ากับโครงสร้าง ρ_k และ ϕ_{kk} ของ ARIMA นั้นมากที่สุด

ด้วยวิธีการเลือกตัวแบบโดยการเปรียบเทียบโครงสร้างแปรผันของค่าตัวอย่าง $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ กับโครงสร้างแปรผันของค่าทางทฤษฎี ρ_k และ ϕ_{kk} จึงจำเป็นที่นักพยากรณ์จะต้องทราบลักษณะโครงสร้างเชิงทฤษฎีของ ρ_k และ ϕ_{kk} ของกระบวนการ AR, MA และ ARIMA ที่อันดับต่าง ๆ เพื่อจะได้เลือกกระบวนการและอันดับ เป็นตัวแบบ ARIMA ทดลอง สำหรับข้อมูลอนุกรมเวลาที่พิจารณา

สรุปลักษณะแปรผันของ ACF และ PACF ของกระบวนการอนุกรมเวลาคงที่สำหรับกระบวนการพื้นฐานดังนี้

รูปแบบ	ACF	PACF
AR(1)	ค่า ρ_k ลดลงอย่างรวดเร็วขณะที่ k >1	ค่า ϕ_{kk} จะมีค่าสูงที่ k = 1 และเท่ากับ 0 เมื่อ k >1
AR(2)	ค่า ρ_k ลดลงอย่างรวดเร็วขณะที่ k >1	ค่า ϕ_{kk} จะมีค่าสูงที่ k = 1,2 และเท่ากับ 0 เมื่อ k >2
MA(1)	ค่า ρ_k จะมีค่าสูงที่ k = 1 และเท่ากับ 0 เมื่อ k >1	ค่า ϕ_{kk} ลดลงอย่างรวดเร็วขณะที่ k>1
MA(2)	ค่า ρ_k จะมีค่าสูงที่ k =1,2 และเท่ากับ 0 เมื่อ k>2	ค่า ϕ_{kk} ลดลงอย่างรวดเร็วขณะที่ k>1
ARMA(1,1)	ค่า ρ_k ลดลงอย่างรวดเร็วหลังจากแล็ก k = 1	ค่า ϕ_{kk} ลดลงอย่างรวดเร็วหลังแล็ก k = 1

สำหรับกระบวนการ AR, MA และ ARIMA ที่อันดับอื่น ๆ พิจารณาได้ในทำนองเดียว

ในองค์ประกอบที่เป็นฤดูกาลมีคาบเวลา s การกำหนดอันดับ P และ Q พิจารณาทำนองเดียวกันกับองค์ประกอบที่ไม่เป็นฤดูกาล โดยพิจารณาโครงสร้างแปรผันของอัตสหสัมพันธ์ $\hat{\rho}_k$

และ $\hat{\phi}_{kk}$ ที่แล็ก ฤดูกาล $s,2s,3s,\ldots$ เปรียบเทียบกับโครงสร้างของ ρ_k และ ϕ_{kk} ทางทฤษฎี ซึ่งมีลักษณะตามที่กล่าวไปแล้วข้างต้น

4.2 การประมาณค่าพารามิเตอร์ในตัวแบบ ARIMA

เมื่อนักพยากรณ์เลือกตัวแบบ ARIMA ทดลองได้แล้ว ขั้นต่อไปคือประมาณค่าของพารามิเตอร์ที่ปรากฏในตัวแบบ วิธีการประมาณค่าพารามิเตอร์มีด้วยกันหลายวิธี วิธีหนึ่งที่ใช้กันมากคือ วิธีกำลังสองน้อยที่สุดแบบไม่เป็นเชิงเส้น(Nonlinear least-squares method)

เมื่อใช้โปรแกรมคอมพิวเตอร์ในเรื่อง ARIMA ในขั้นประมาณค่าพารามิเตอร์นอกจากจะได้ค่าประมาณของพารามิเตอร์แล้ว จะมีค่าของตัวสถิติต่าง ๆปรากฏออกมาด้วยซึ่งจะใช้ประโยชน์ในการทดสอบทางสถิติว่า องค์ประกอบหรือพารามิเตอร์นั้นควรมีอยู่ในตัวแบบหรือไม่ซึ่งนั่นคือเป็นหนทางหนึ่งในการพิจารณาว่า ตัวแบบที่พิจารณานั้นเหมาะสมเพียงพอหรือไม่ในทางสถิติ

4.3 การวินิจฉัยตัวแบบ ARIMA

ภายหลังที่ประมาณค่าพารามิเตอร์ของตัวแบบ ARIMA แล้ว นักพยากรณ์ควรจะตรวจสอบตัวแบบก่อนที่จะตัดสินใจนำตัวแบบนั้นไปใช้พยากรณ์ เนื่องจากตัวแบบที่พิจารณาคัดเลือกในขั้นแรกนั้นอาจยังเลือกไม่ถูกต้องเหมาะสม จึงควรวินิจฉัย และถ้าพบว่ายังไม่เหมาะสมควรกลับไปขั้นที่ 1 พิจารณาปรับปรุงแก้ไขตัวแบบใหม่ และดำเนินการขั้นที่ 2 ประมาณค่าและวินิจฉัยในขั้นที่ 3 กรรมวิธีกระทำซ้ำ ๆ เช่นนี้จนกว่าจะได้ตัวแบบที่เหมาะสมเพียงพอในเชิงสถิติ

การวินิจฉัยตัวแบบจะกระทำการตรวจสอบคุณสมบัติเชิงสถิติของค่าผิดพลาดสุ่ม (a_t) และการทดสอบว่า ค่าผิดพลาดสุ่มมีอัตสัมพันธ์หรือไม่ จะเป็นการตรวจสอบที่สำคัญ มากที่สุดในการวินิจฉัยความเพียงพอในเชิงสถิติของตัวแบบ ARIMA ฉะนั้นการตรวจสอบจะคำนวณค่า SACF และSPACF ของค่าเศษเหลือตกค้าง $e_t = y_t - \bar{y}_t$ ซึ่งเป็นค่าประมาณของ a_t ที่แล็ก k ต่าง ๆ และทดสอบด้วยค่าของตัวสถิติที่ t สำหรับทดสอบว่าค่าผิดพลาดสุ่มมีอัตสหสัมพันธ์หรือไม่ที่แต่ละแล็ก $k = 1, 2, 3, \ldots, m$ และทดสอบอัตสหสัมพันธ์รวมหรือพร้อมกัน k แล็ค ด้วยตัวสถิติไคกำลังสอง (Chi-squared test) ว่าค่าผิดพลาดไม่มีอัตสหสัมพันธ์ k แล็คแรกนอกจากการวินิจฉัยตัวแบบด้วยการทดสอบเชิงสถิติแล้ว นักพยากรณ์อาจดำเนินการตรวจสอบด้วยวิธีอื่นๆ ด้วยเช่นกัน การ

เขียนกราฟของเศษเหลือตกค้างกับแกนเวลา ถ้าพบว่าค่าของเศษเหลือตกค้างกระจายเป็นแนวโน้มลักษณะขนานรอบค่าเฉลี่ยศูนย์แสดงเหตุผลได้ว่า ค่าผิดพลาดมีค่าเฉลี่ยศูนย์ และมีค่าแปรปรวนคงที่ แต่ถ้าการกระจายของค่าเศษเหลือตกค้างมีรูปแบบต่างไปจากแนวขนาน ควรพิจารณาปรับปรุงแก้ไขตัวแบบซึ่งอาจจะพบว่าความแปรปรวนยังไม่คงที่ต้องปรับให้คงที่ด้วยวิธีการแปลงข้อมูล เป็นต้น

ผลจากการวินิจฉัยตัวแบบนอกจากจะช่วยตรวจสอบว่าตัวแบบที่กำลังพิจารณาเหมาะสมเพียงพอในเชิงสถิติหรือไม่แล้ว ยังมีแนวทางในการปรับปรุงแก้ไขตัวแบบด้วยถ้าพบว่าตัวแบบยังไม่เหมาะสม กล่าวคือจากลักษณะของ SACF และ SPACF ของเศษเหลือตกค้างอาจพบว่าควรเพิ่มองค์ประกอบ MA เข้าในตัวแบบ ถ้ายังไม่มีองค์ประกอบ MA หรืออันดับของ MA ในตัวแบบให้มากขึ้น หรืออาจพบว่าควรเพิ่มองค์ประกอบ AR หรืออันดับของ AR ในตัวแบบเป็นต้น

2.2.2 ทฤษฎีวิธีการปรับให้เรียบ

วิธีการปรับให้เรียบสำหรับการพยากรณ์ ข้อมูลอนุกรมเวลามีหลายวิธีด้วยกัน วิธีการเหล่านี้โดยทั่วไปเป็นวิธีการที่ไม่ซับซ้อนใช้เวลาไม่มากในการ คำนวณและสำหรับความถูกต้องของค่าพยากรณ์จะมากน้อยขึ้นอยู่กับเทคนิคพยากรณ์ที่เลือกใช้สอดคล้องกับลักษณะข้อมูลอนุกรมเวลาเพียงใดส่วนการทำการวิจัยครั้งนี้ใช้วิธีการปรับให้เรียบครั้งเดียวแบบเอกซโพแนนเชียล (Single Exponential Smoothing Method) วิธีการปรับให้เรียบสองครั้งแบบเอกซโพเนนเชียล (Double Exponential Smoothing Method) วิธีการพยากรณ์ของโฮลท์ (Holt's Forecast Method) วิธีการพยากรณ์ของวินเตอร์ (Winter's Forcast Method)

2.2.2.1 การปรับให้เรียบครั้งเดียวแบบเลขชี้กำลัง(Single Exponential Smoothing)

วิธีพยากรณ์วิธีหนึ่งสำหรับอนุกรมเวลาที่มีค่าเฉลี่ยเปลี่ยนแปลงน้อย ๆ เป็นการหาค่าเฉลี่ยแบบถ่วงน้ำหนัก คือวิธีการปรับให้เรียบครั้งเดียวแบบเลขชี้กำลัง วิธีนี้จะให้น้ำหนักกับข้อมูลปัจจุบันมากที่สุด และให้น้ำหนักลดลงเรื่อย ๆ สำหรับข้อมูลอดีตตามลำดับ ซึ่งเมื่อเขียนกับข้อมูลปัจจุบันมากที่สุด และให้น้ำหนักลดลงเรื่อย ๆ สำหรับข้อมูลอดีตตามลำดับ ซึ่งเมื่อเขียนกราฟแสดงการลดลงของน้ำหนักจะมีรูปแบบเลขชี้กำลัง และวิธีนี้เหมาะสมสำหรับการพยากรณ์ในระยะสั้น

ตัวแบบพยากรณ์ค่าในอนาคต Y_{t+l} (l = 1,2,...) จากเวลาปัจจุบัน t :

$$\hat{Y}_t(l) \quad = \quad S_t \quad = \quad \alpha Y_t + (1-\alpha)S_{t-1} \qquad ;t = 1,2,..$$

ซึ่ง α คือ ค่าคงที่หรือสัมประสิทธิ์ปรับให้เรียบ มีค่าอยู่ระหว่าง 0 และ 1 นักพยากรณ์จะต้องเลือกกำหนดค่า α ซึ่งโดยทั่วไปจะเลือกค่า α อยู่ระหว่าง 0.05 และ 0.3 ซึ่งถ้าระดับค่าเฉลี่ยเปลี่ยนแปลงช้า ๆ ควรเลือก ค่าเล็ก สำหรับ S_t , มีชื่อเรียกว่า "ตัวสถิติปรับให้เรียบ" ซึ่งการคำนวณหาค่า S_t จะต้องใช้ค่า S_{t-1} , S_{t-2} , ..., S_1, S_0 นั่นคือ เมื่อทราบค่าเริ่มต้น S_0 จะสามารถหาค่า S ตัวต่อ ๆ ไปได้ ฉะนั้นนักพยากรณ์จะต้องกำหนดค่าเริ่มต้น S_0 และตัวอย่างการเลือกกำหนดค่า S_0 เช่น ใช้ค่าเฉลี่ยของอนุกรมเวลาที่มีอยู่

$$S_0 \quad = \quad \bar{Y} \quad = \frac{1}{t}(Y_1 + Y_2 + ... + Y_t)$$

โดยเฉพาะกรณีที่ระดับค่าเฉลี่ยของอนุกรมเวลาเปลี่ยนแปลงช้า ๆ

อีกหนทางหนึ่งในการเลือกค่า α คือทดลองแปรเปลี่ยนค่า α เช่น α เริ่มจาก α=0.01 ต่อไปเป็น 0.02, 0.03, ... และแต่ละค่า α คำนวณค่า S_t และหาค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสอง(MSE) จากนั้นเปรียบเทียบค่า MSE ทั้งหมดและเลือกค่า α ที่ให้ MSE ต่ำสุด

2.2.2.2 **วิธีการปรับให้เรียบสองครั้งแบบเลขชี้กำลัง**(Double Exponential Smoothing Method)

จากแนวคิดการปรับให้เรียบครั้งเดียวแบบเลขชี้กำลัง นำมาขยายผลใช้กับข้อมูลอนุกรมเวลาที่มีแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วงเวลา t มีสูตรพยากรณ์ค่าจริง Y_{t+l} ที่เวลา $t+l$ จากเวลาปัจจุบัน t ดังนี้

$$\hat{Y}_t(l) \quad = \quad \left(2 + \frac{\alpha l}{1-\alpha}\right)S_t^{[1]} - \left(1 + \frac{\alpha l}{1-\alpha}\right)S_t^{[2]}$$

ซึ่ง

$$S_t^{[1]} \quad = \quad \alpha Y_t + (1-\alpha)S_{t-1}^{[1]}$$

$$S_t^{[2]} \quad = \quad \alpha Y_t^{[1]} + (1-\alpha)S_{t-1}^{[2]}$$

การคำนวณ $S_t^{[1]}$ และ $S_t^{[2]}$ ต้องการทราบค่า $S_{t-2}^{[1]}$, $S_{t-2}^{[2]}$, $S_{t-3}^{[1]}$,..., $S_0^{[1]}$, $S_0^{[2]}$ และดังนั้นต้องเริ่มด้วยค่า $S_0^{[1]}$ และ $S_0^{[2]}$ เราประมาณค่า $S_0^{[1]}$ และ $S_0^{[2]}$ ได้ดังนี้

$$S_0^{[1]} = \hat{\beta}_0 - \left(\frac{1-\alpha}{\alpha}\right)\hat{\beta}_1$$

$$S_0^{[2]} = \hat{\beta}_0 - 2\left(\frac{1-\alpha}{\alpha}\right)\hat{\beta}_1$$

ซึ่ง

$$\hat{\beta}_0 = \bar{Y} - \left(\frac{\bar{T}+1}{2}\right)\hat{\beta}_1$$

$$\hat{\beta}_1 = 12\sum_{t=1}^{T}\frac{\left(t-(T+1)/2\right)Y_t}{T^3-T}$$

2.2.2.3 วิธีการของโฮลท์ (Holt's Forecast Method)

วิธีการของโฮลท์มีลักษณะคล้ายกับวิธีการปรับให้เรียบสองครั้งแบบเอกซโพเนนเชียลแต่มีลักษณะทั่วไปมากกว่า มีสูตรพยากรณ์ดังนี้

$$\hat{Y}_t(l) = S_t + l\hat{\beta}_t$$

ซึ่ง

ตัวสถิติปรับระดับ $S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + \hat{\beta}_{t-1})$

ตัวสถิติปรับแนวโน้ม $\hat{\beta}_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\hat{\beta}_{t-1}$

เห็นได้ว่าวิธีการของโฮลท์ใช้พารามิเตอร์ปรับให้เรียบสองตัวคือ α ($0<\alpha<1$) และ γ ($0<\gamma<1$) ซึ่งนักพยากรณ์จะต้องกำหนดค่าทั้งสองนี้ ในการกำหนดผู้วิเคราะห์สามารถกำหนดเองหรืออาจให้โปรแกรมสำเร็จรูปสำหรับการพยากรณ์ค้นหาให้เพื่อให้การพยากรณ์มีความคาดเคลื่อนน้อยที่สุดและผู้วิเคราะห์ต้องกำหนดค่าเริ่มต้น S_t และ $\hat{\beta}_1$

2.2.2.4 **วิธีการของวินเตอร์** (Winter's Forecast Method)

เมื่ออนุกรมเวลามีองค์ประกอบฤดูกาล การใช้วิธีการพยากรณ์ที่กล่าวมาทั้งหมดสำหรับการพยากรณ์จะไม่เหมาะสม ควรจะใช้วิธีการพยากรณ์ที่พิจารณาองค์ประกอบฤดูกาลร่วมด้วย วิธีการพยากรณ์ของวินเตอร์ เป็นการขยายผลของวิธีการพารามิเตอร์สองตัวของโฮลท์โดยเพิ่มพารามิเตอร์หรือค่าคงที่อีกหนึ่งตัวรวมเป็นสามตัวคือ ค่าคงที่ปรับให้เรียบระดับ (α) ค่าคงที่ปรับให้เรียบสำหรับแนวโน้มหรือความชัน (γ) และค่าคงที่ปรับให้เรียบสำหรับฤดูกาล (δ) ค่าทั้งสามมีค่าอยู่ระหว่าง 0 และ 1

ตัวแบบอนุกรมเวลาตัวแบบหนึ่งของวินเตอร์ ซึ่งประกอบด้วยองค์ประกอบแนวโน้มและฤดูกาลมีสมการดังนี้ และมีชื่อเรียกว่าตัวแบบผลคูณของวินเตอร์

$$Y_t = (\mu_t + \beta_t t) I_t + \varepsilon_t$$

ซึ่ง μ_t , β_t , I_t เป็นพารามิเตอร์แสดงระดับ ความชัน และฤดูกาล ของอนุกรมเวลาตามลำดับ และ ε_t คือความคลาดเคลื่อนสุ่ม โดยมีข้อมูลสมมติพื้นฐานคือ มีค่าเฉลี่ยศูนย์ ความแปรปรวนคงที่และไม่มีสหสัมพันธ์

ตัวแบบข้างต้นเหมาะสมกับอนุกรมเวลามีการแกว่งหรือการผันแปรของฤดูกาลเป็นสัดส่วนโดยตรงกับระดับของอนุกรม (ค่าเฉลี่ยของอนุกรม) กล่าวคือ การแกว่งจะมากขึ้นขณะที่ระดับของอนุกรม เพิ่มขึ้นส่วนการผันแปรของ ε_t ไม่ขึ้นอยู่กับระดับของอนุกรม

จากตัวแบบข้างต้น ได้สูตรพยากรณ์ดังนี้

$$\hat{Y}_t(l) = \left(\hat{\mu}_t(l) + \hat{\beta}_t l\right)\hat{I}_{t+l-m}, \quad t = m, m+1, \ldots$$

ซึ่ง

$$\hat{\mu}_t = \alpha\left(Y_t / \hat{I}_{t-m}\right) + (1-\alpha)\left(\hat{\mu}_{t-1} + \hat{\beta}_{t-1}\right)$$

$$\hat{\beta}_t = \gamma\left(\hat{\mu}_t - \hat{\mu}_{t-1}\right) + (1-\gamma)\hat{\beta}_{t-1}$$

$$\hat{I}_t = \delta\left(Y_t / \hat{\mu}_t\right) + (1-\delta)\hat{I}_{t-m}$$

m = ความยาวของคาบฤดูกาล

การคำนวณค่าพยากรณ์ $\hat{Y}_t(l)$ ต้องกำหนดค่าเริ่มต้นของ $\hat{\mu}_t, \hat{\beta}_t$ และ $\hat{I}_t$ นอกเหนือจากการกำหนดค่าคงที่ปรับให้เรียบ α , γ, δ และหนทางหนึ่งในการกำหนดค่าเริ่มต้นคือให้

$$\begin{aligned} \hat{\mu}_m &= (Y_1 + Y_2 + \ldots + Y_m)/m \\ \hat{I}_t &= Y_t / \hat{\mu}_m \qquad t = 1,2,\ldots,m \\ \hat{\beta}_m &= 0 \end{aligned}$$

2.2.3 การวิเคราะห์อนุกรมเวลาแบบคลาสสิค (Classical Decomposition Method)

อนุกรมเวลา $\{Y_t\}$ จะมีการเคลื่อนไหวตามกาลเวลาในรูปแบบต่างๆ เช่นมีแนวโน้ม และมีฤดูกาล ดังนั้นหนทางหนึ่งในการวิเคราะห์อนุกรมเวลา เพื่อหาตัวแบบพยากรณ์ คือ วิเคราะห์หาองค์ประกอบในข้อมูลอนุกรมเวลาที่ศึกษา มาสร้างเป็นตัวแบบพยากรณ์ คือ วิเคราะห์หาองค์ประกอบในข้อมูลอนุกรมเวลาที่ศึกษา มาสร้างเป็นตัวแบบพยากรณ์

องค์ประกอบของอนุกรมเวลา จำแนกได้เป็น 3 องค์ประกอบหลัก คือ องค์ประกอบแนวโน้ม - วัฏจักร(trend - cycle component) องค์ประกอบฤดูกาล(seasonal component) และองค์ประกอบไม่ปกติ หรือองค์ประกอบส่วนเหลือ(irregular or remainder component) และดังนั้นโดยวิธีแยกองค์ประกอบจะได้ตัวแบบสำหรับการพยากรณ์ค่า Y_t และเรียกตัวแบบที่ได้ว่า ตัวแบบแยกองค์ประกอบ ซึ่งจำแนกได้เป็น 2 ตัวแบบคือ

2.2.3.1 ตัวแบบเชิงบวก (Additive Model)

$$Y_t = T_t + S_t + \varepsilon_t$$

2.2.3.2 ตัวแบบเชิงคูณ (Multiplicative Model)

$$Y_t = T_t * S_t * \varepsilon_t$$

โดยที่

Y_t คือ ค่าของอนุกรมเวลา ณ คาบเวลา t

ε_t คือ ค่าคลาดเคลื่อนสุ่ม

T_t คือ องค์ประกอบแนวโน้ม-วัฏจักร ณ คาบเวลา t ซึ่งโดยทั่วไปจะอยู่ในรูปแบบพหุนามอันดับต่ำ (low-order polynomial) เช่น $T_t = \beta_0 + \beta_1 t$ และ $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$ เป็นต้น

S_t คือ องค์ประกอบฤดูกาล ณ คาบเวลา t ซึ่งอาจจะอยู่ในรูปแบบของตัวแปรบ่งชี้ฤดูกาล (seasonal component) เช่น

$S_t = \alpha_1 I_{1,t} + \alpha_2 I_{2,t} + ... + \alpha_{11} I_{11,t}$ ($I_{1,t}$ เป็นตัวแปรบ่งชี้ฤดูกาล) หรืออยู่ในรูปแบบฟังก์ชันตรีโกณ เช่น

$$S_t = \phi_1 \sin\left(\frac{2\pi t}{12}\right) + \phi_2 \cos\left(\frac{2\pi t}{12}\right) + \phi_3 \sin\left(\frac{4\pi t}{12}\right) + \phi_4 \cos\left(\frac{4\pi t}{12}\right)$$

ตัวแบบเชิงบวก เหมาะสมสำหรับข้อมูลอนุกรมเวลาที่มีความแปรปรวนคงที่ ไม่แปรผันตามเวลา หรือการแกว่งของฤดูกาลไม่แปรผันตามระดับของอนุกรมเวลา มิฉะนั้น ตัวแบบเชิงคูณจะเหมาะสมกว่า อย่างไรก็ตาม เราสามารถใช้วิธีการแปลงข้อมูลอนุกรมเวลาที่ไม่คงที่ในความแปรปรวน หรือแปลงตัวแบบเชิงคูณเป็นตัวแบบเชิงบวกได้ โดยการใส่ ***ln*** ซึ่งจะได้ตัวแบบ

$$\ln Y_t = \ln T_t + \ln S_t + \ln \varepsilon_t$$

นั่นคือ ใช้ตัวแบบเชิงบวกกับข้อมูลที่ใส่ ***ln***

ตัวอย่างตัวแบบพยากรณ์แบบแยกองค์ประกอบ :

$$Y_t = b_0 + b_1 t + b_2 I_{1,t} + b_3 I_{2,t} + ... + b_{12} I_{11,t}$$

โดยที่ t แทนคาบเวลา

$I_{i,t}$ (i = 1,2,...,11) แทนตัวบ่งชี้ สำหรับเดือนที่ i ในคาบเวลา t (ให้ $I_{i,t} = 1$ สำหรับเดือนที่ i ในคาบเวลา t มิฉะนั้น $I_{i,t} = 0$)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.2.4 การหาค่าพยากรณ์ร่วม (Combined Forecasts)

$$CF = \sum_{j=1}^{m} W_j \hat{Y}_{jt}$$

โดยที่ W_j คือ ค่าเฉลี่ยถ่วงน้ำหนักของวิธีการพยากรณ์ที่ j

$\hat{Y}_{jt}$ คือ ค่าพยากรณ์ของวิธีการพยากรณที่ j ที่คาบเวลาที่ t

j คือ วิธีการพยากรณ์ที่ j ; j = 1,2,...,m

t คือ คาบเวลา ; t = 1,2,...,n

2.2.4.1 วิธีการให้น้ำหนักที่เท่ากัน (Simple Average Method)

วิธีการนี้ให้ความสำคัญแก่วิธีการพยากรณ์ที่นำมารวมกัน ด้วยน้ำหนักที่เท่ากัน มีสูตรในการคำนวณดังนี้

$$W_j = \frac{1}{m}$$

โดยที่ W_j คือ ค่าเฉลี่ยน้ำหนักของวิธีการพยากรณ์ที่ j

m คือ จำนวนวิธีการพยากรณ์ที่นำมารวมกัน

2.2.4.2 วิธีค่าสัมบูรณ์ต่ำ สุด (Least Absolute Value Method)

เป็นวิธีการที่คำนวณหาค่าเฉลี่ยถ่วงน้ำหนัก ซึ่งอาศัยเทคนิคสมการเชิงเส้นตรง (Linear Programming Technique) โดยมีหลักการทำให้ผลรวมของค่าสัมบูรณ์ของค่าความคลาดเคลื่อนมีค่าต่ำสุด มีสูตรการคำนวณดังสมการต่อไปนี้

$$\text{หาค่าต่ำสุด } \sum|\varepsilon_t| \qquad (1)$$

โดยที่

$$\varepsilon_t = \frac{Y_t - CF}{Y_t} \qquad , t = 1,2,...,n$$

จากสมการที่ (1) แปลงให้เป็นรูปแบบเชิงเส้นดังนี้

กำหนดให้

$$Z = \sum|\varepsilon_t|$$

เนื่องจาก ε_t ไม่มีข้อจำกัดด้านเครื่องหมาย จะได้ว่า

$$\varepsilon_t = \varepsilon_t^+ - \varepsilon_t^- \qquad ; \varepsilon_t^+ , \varepsilon_t^- \geq 0 \qquad (2)$$

โดยนิยาม

$$\varepsilon_t^+ = \begin{cases} \varepsilon_t & ; \varepsilon_t \geq 0 \\ 0 & ; \varepsilon_t < 0 \end{cases}$$

$$\varepsilon_t^- = \begin{cases} 0 & ; \varepsilon_t \geq 0 \\ -\varepsilon_t & ; \varepsilon_t < 0 \end{cases}$$

เพราะฉะนั้น ได้ว่า $\varepsilon_t^+ \times \varepsilon_t^- = 0$

นั้นคืออย่างน้อย 1 ตัวแปรใน ε_t^+ และ ε_t^- เท่ากับ 0 เสมอ
ดังนั้น

$$|\varepsilon_t| = \varepsilon_t^+ + \varepsilon_t^-$$

และให้ $U_{t-1} = \varepsilon_t^+$ และ $V_{t-1} = \varepsilon_t^-$
ดังนั้นแปลงปัญหา (1) ได้เป็น

$$\text{หาค่าต่ำสุด } \sum (U_{t-1} + V_{t-1})$$

โดยมีเงื่อนไขคือ

$$\frac{Y_t - \sum_{j=1}^{m} W_j \hat{Y}_{jt}}{Y_t} = U_{t-1} - V_{t-1} \qquad , t = 2,3,\ldots,n$$

และ

$$\sum_{j=1}^{m} W_j = 1$$

$$W_j , U_{t-1} , V_{t-1} \geq 0$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างปัญหาข้างต้น เมื่อ m = 4 และ n = 5 เขียนแสดงรูปแบบเต็มได้ดังนี้

หาค่าต่ำสุด

$$Z = 0W_1 + 0W_2 + 0W_3 + 0W_4 + U_1 + U_2 + U_3 + U_4 + V_1 + V_2 + V_3 + V_4$$

เงื่อนไข :

$$\frac{W_1\hat{Y}_{12}}{Y_2} + \frac{W_2\hat{Y}_{22}}{Y_2} + \frac{W_3\hat{Y}_{32}}{Y_2} + \frac{W_4\hat{Y}_{42}}{Y_2} + U_1 - V_1 = 1$$

$$\frac{W_1\hat{Y}_{13}}{Y_3} + \frac{W_2\hat{Y}_{23}}{Y_3} + \frac{W_3\hat{Y}_{33}}{Y_3} + \frac{W_4\hat{Y}_{43}}{Y_3} + U_2 - V_2 = 1$$

$$\frac{W_1\hat{Y}_{14}}{Y_4} + \frac{W_2\hat{Y}_{24}}{Y_4} + \frac{W_3\hat{Y}_{34}}{Y_4} + \frac{W_4\hat{Y}_{44}}{Y_4} + U_3 - V_3 = 1$$

$$\frac{W_1\hat{Y}_{15}}{Y_5} + \frac{W_2\hat{Y}_{25}}{Y_5} + \frac{W_3\hat{Y}_{35}}{Y_5} + \frac{W_4\hat{Y}_{45}}{Y_5} + U_4 - V_4 = 1$$

$$W_1 + W_2 + W_3 + W_4 = 1$$

$$W_j, U_{i-1}, V_{i-1} \geq 0$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย