บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการวิจัยทั้งในสาขาวิทยาศาสตร์และสังคมศาสตร์ เช่น การวิจัยทางสาธารณสุข การวิจัยทางเศรษฐศาสตร์ การวิจัยทางการศึกษา และการวิจัยทางพฤติกรรมศาสตร์ ฯลฯ จำเป็นต้องอาศัยวิธีการทางสถิติในการค้นคว้าหาคำตอบและพยากรณ์ และต้องมีการดำเนินงานวิจัยอย่างมีระบบเพื่อให้งานวิจัยที่ได้มีความถูกต้องและมีความน่าเชื่อถือ การวิเคราะห์ความถดถอย (regression analysis) เป็นวิธีการทางสถิติที่ถูกนำมาใช้กันอย่างแพร่หลายในปัจจุบัน โดยต้องการศึกษาความสัมพันธ์ระหว่างตัวแปรที่เราต้องการศึกษา (ตัวแปรตาม) กับตัวแปรตัวอื่นๆที่เกี่ยวข้อง (ตัวแปรอิสระ) ซึ่งถ้ามีตัวแปรอิสระเพียงตัวเดียวจะเรียกว่า การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (simple linear regression) แต่ถ้ามีตัวแปรอิสระมากกว่าหนึ่งตัวจะเรียกว่าการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (multiple linear regression) และมีรูปแบบดังนี้

(1.1) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนสุ่มขนาด $n \times 1$

n เป็นขนาดตัวอย่าง

และ p เป็นจำนวนตัวแปรอิสระ

เนื่องจากการพิจารณาตัวแปรอิสระที่ใช้ในตัวแบบถดถอยนั้นอาจพบว่ามีตัวแปรอิสระหลายตัวแปรที่มีอิทธิพลต่อตัวแปรตาม ทำให้ต้องใช้ตัวแปรอิสระหลายตัวแปรในตัวแบบถดถอย ดังนั้นการคัดเลือกตัวแปรอิสระจึงเป็นสิ่งที่สำคัญในการวิเคราะห์ความถดถอยเชิงเส้นเพื่อนำไปสู่การคัดเลือกตัวแบบถดถอยที่ดีที่สุดซึ่งจะทำให้ค่าพยากรณ์มีความถูกต้องและแม่นยำ การคัดเลือกตัวแบบที่ดีที่สุดเพื่อใช้ในการพยากรณ์จึงจำแนกเป็น 2 แนวทาง[1] คือ

1. ใช้ค่าสถิติในการพิจารณาความเหมาะสมของตัวแบบทุกรูปแบบที่เป็นไปได้หรือเรียกว่า

---

[1] ทรงศิริ แต้สมบัติ, การวิเคราะห์การถดถอย (กรุงเทพมหานคร : มหาวิทยาลัยเกษตรศาสตร์, 2541), หน้า 205.

วิธีพิจารณาทุกรูปแบบ (all possible regression) ค่าสถิติที่นิยมใช้ในวิธีนี้ เช่น ค่าสัมประสิทธิ์การกำหนด (R^2) ค่าสัมประสิทธิ์การกำหนดที่ปรับแล้ว (R^2_{adj}) และค่าสถิติของมอลโลวส์ ($Mallow\, C_p$) เป็นต้น

2. ใช้วิธีการคัดเลือกตัวแปรอิสระ (variable selection) โดยการเพิ่ม และ/หรือ ลดตัวแปรอิสระจากตัวแบบถดถอย ซึ่งวิธีที่นิยมใช้กันอย่างแพร่หลาย เช่น วิธีการคัดเลือกตัวแปรแบบไปข้างหน้า (forward selection) วิธีการกำจัดตัวแปรแบบถอยหลัง (backward elimination) และวิธีการถดถอยแบบขั้นบันได (stepwise regression) เป็นต้น

การใช้วิธีการในการคัดเลือกตัวแบบถดถอยจาก 2 แนวทางข้างต้นอาจทำให้เกิดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ได้ เพราะวิธีการเหล่านี้ใช้เพียงข้อมูลในปัจจุบันโดยไม่ได้คำนึงถึงข้อมูลในอดีต ดังนั้นการนำข้อมูลในอดีตมาพิจารณาด้วยจะช่วยลดค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ได้ ซึ่งแนวคิดของเบส์ที่นำมาใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นจะเรียกว่า การวิเคราะห์เชิงเบส์ (Bayesian approach) โดยใช้การแจกแจงก่อน (prior distribution) ของพารามิเตอร์มาพิจารณาด้วย การวิเคราะห์เชิงเบส์จึงเป็นสิ่งที่นักสถิติให้ความสนใจมากในปัจจุบัน

[2]ในปี ค.ศ. 1997 ราฟเทอร์รี (Raftery) เมดิแกน (Madigan) และโฮเอ็ททิง (Hoeting) ได้เสนอวิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) สำหรับตัวแบบการถดถอยเชิงเส้น ซึ่งวิธีนี้คำนึงถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบ (model uncertainty) วิธีการเฉลี่ยตัวแบบของเบส์จะพิจารณาค่าความน่าจะเป็นภายหลัง (posterior probability) สำหรับทุกๆ ตัวแบบที่เราสนใจและนำตัวแบบทุกตัวแบบที่เราสนใจมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนักเพื่อหาค่าพยากรณ์ที่เหมาะสม การหาปริภูมิตัวแบบที่จะนำมาใช้ในวิธีการเฉลี่ยตัวแบบของเบส์นั้น ราฟเทอร์รี เมดิแกน และโฮเอ็ททิง ได้เสนอไว้ 2 แนวทางดังนี้คือ

1. วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคมวินโดว์ (Occam's Window) (BMA_{OCC})

2. วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC^3})

---

[2]จิตติมา ผสมญาติ, การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2546), หน้า 2.

[3]ในปี ค.ศ. 2002 บาร์บีรี (Barbieri) และเบอร์เกอร์ (Berger) ได้เสนอวิธีการคัดเลือกตัวแบบของเบส์สำหรับการคัดเลือกระหว่างกลุ่มของตัวแบบเชิงเส้นปกติ (selection among normal linear models) ตัวแบบที่เหมาะสมที่สุดอาจไม่ใช่ตัวแบบที่มีค่าความน่าจะเป็นภายหลังสูงสุด (the model with highest posterior probability) แต่กลับเป็นตัวแบบที่มีค่าความน่าจะเป็นภายหลังกลางๆ (the median probability model) และเกณฑ์ที่ใช้ในการเลือกตัวแบบที่เหมาะสมที่สุดจากกลุ่มของตัวแบบที่มีความน่าจะเป็นภายหลังกลาง ๆ คือความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสอง (square error loss) ซึ่งจะเลือกตัวแบบที่มีค่าความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสองต่ำสุด

ในปี ค.ศ. 2002 นิทัสน์ สุขสุวรรณ เปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ โดยจะเปรียบเทียบวิธีการคัดเลือกสมการถดถอย 5 วิธี ได้แก่ วิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์ (BIC) วิธีการคัดเลือกตัวแปรของเบส์ (BVS) วิธีการเฉลี่ยตัวแบบของเบส์โดยใช้การค้นหาปริภูมิตัวแบบ ด้วยวิธีออกส์แคมวินโดว์ (BMA_{OCC}) วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (BMA_{MC^3}) และวิธีการถดถอยแบบขั้นบันได (SR) เกณฑ์ที่ใช้ในการตัดสินใจคือเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) ซึ่งผลการวิจัยได้ข้อสรุปดังนี้ การเปรียบเทียบค่า AMSE ของทั้ง 5 วิธีเรียงลำดับจากน้อยไปมาก ได้แก่ วิธี BMA_{MC^3} BMA_{OCC} BVS BIC และ SR ตามลำดับ สำหรับทุกสถานการณ์ วิธี BMA_{OCC} จะให้ค่า AMSE สูงกว่าวิธี BMA_{MC^3} เพียงเล็กน้อย โดยเฉพาะอย่างยิ่งเมื่อระดับนัยสำคัญลดลง สำหรับวิธี BVS จะให้ค่า AMSE ใกล้เคียงกับวิธี BMA_{MC^3} และวิธี BMA_{OCC} ก็ต่อเมื่อค่าคงที่ $(\sigma_\beta/\tau,c)$ มีค่าต่ำๆ ส่วนวิธี BIC และวิธี SR มีค่า AMSE แตกต่างจากวิธี BMA_{MC^3} และวิธี BMA_{OCC} อย่างชัดเจนในทุกสถานการณ์ ปัจจัยที่มีผลต่อค่า AMSE ของทุกวิธี คือ ขนาดตัวอย่าง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ โดยที่ค่า AMSE จะแปรผันตามค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ แต่จะแปรผกผันกับขนาดตัวอย่าง นอกจากนั้นค่า AMSE ของ 2 วิธี คือ วิธี BMA_{OCC} และวิธี SR จะแปรผันตามระดับนัยสำคัญ โดยที่ 3 วิธีที่เหลือค่า AMSE ไม่เปลี่ยนแปลงเพราะไม่ได้นำระดับนัยสำคัญมาพิจารณา ส่วนค่า AMSE ของ 2 วิธี คือวิธี BMA_{MC^3} และวิธี BVS จะแปรผันตามค่าคงที่ $(\sigma_\beta/\tau,c)$ โดยที่ 3 วิธีที่เหลือค่า AMSE ไม่เปลี่ยนแปลงเพราะไม่ได้นำค่าคงที่ $(\sigma_\beta/\tau,c)$ มาพิจารณา

---

[3]จิตติมา ผสมญาติ, การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2546), หน้า 3.

ในปี ค.ศ. 2003 จิตติมา ผสมญาติ เปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์ เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ โดยเปรียบเทียบวิธีการคัดเลือกสมการถดถอย 3 วิธี ได้แก่ วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โล โดยใช้ลูกโซ่มาร์คอฟ (BMA_{MC^3}) เมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ (BMA_{SVT}) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (OPM) และวิธีการถดถอยแบบขั้นบันได (SR) เกณฑ์ที่ใช้ในการตัดสินใจคือเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) ผลการวิจัยได้ข้อสรุปดังนี้ การเปรียบเทียบค่า AMSE ของทั้ง 3 วิธีเรียงลำดับจากน้อยไปมากได้แก่ วิธี BMA_{SVT} OPM และ SR ตามลำดับสำหรับทุกสถานการณ์ วิธี OPM จะได้ค่า AMSE สูงกว่าวิธี BMA_{SVT} เพียงเล็กน้อยโดยเฉพาะอย่างยิ่งเมื่อค่าคงที่ $(\sigma_\beta/\tau, c)$ มีค่าต่ำๆ ส่วนวิธี SR มีค่า AMSE แตกต่างจากวิธี BMA_{SVT} และ OPM อย่างชัดเจนในทุกสถานการณ์ ปัจจัยที่มีผลต่อค่า AMSE ของทุกวิธีคือ ขนาดตัวอย่าง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ โดยที่ค่า AMSE จะแปรผันตามค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ แต่จะแปรผกผันกับขนาดตัวอย่าง นอกจากนั้นค่า AMSE ของ 2 วิธี คือ วิธี BMA_{SVT} และวิธี OPM จะแปรผันตามค่าคงที่ $(\sigma_\beta/\tau, c)$ โดยที่ค่า AMSE ของวิธี SR ไม่เปลี่ยนแปลงเนื่องจากไม่ได้นำค่าคงที่ $(\sigma_\beta/\tau, c)$ มาพิจารณา

ในการวิจัยครั้งนี้ผู้วิจัยได้สนใจทำการศึกษาเปรียบเทียบวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ 3 วิธี คือ

1. วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC^3})

2. วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (Optimal Predictive Model Selection : median probability model (OPM))

3. วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

สองวิธีแรกเป็นวิธีคัดเลือกตัวแบบถดถอยที่ดีที่สุดเชิงเบส์ซึ่งมีการนำข้อมูลในอดีตมาพิจารณาด้วยนั้น ในกรณีที่เราทราบการแจกแจงที่แน่นอนของข้อมูลก็จะทำให้การคัดเลือกตัวแบบถดถอยที่ดีที่สุดเชิงเบส์มีประสิทธิภาพมากขึ้น ในงานวิจัยนี้ได้กำหนดการแจกแจงก่อน(prior distribution) เป็นการแจกแจงแบบแกมมา (gamma distribution) โดยมีการแจกแจงความน่าจะเป็นภายหลัง (posterior distribution) เป็นการแจกแจงแบบแกมมาด้วย ซึ่งเป็นหลักเกณฑ์คู่สังยุคแบบแกมมา (conjugate gamma prior distribution) การใช้เกณฑ์คู่สังยุคแบบแกมมาดังกล่าวเพราะว่าในปัจจุบันได้มีการนำวิธีการทางสถิติมาใช้กับงานวิจัยทางด้านการแพทย์และวิศวกรรม ซึ่งข้อมูลของงานวิจัยเหล่านี้จะพิจารณาอัตราการเกิดเหตุการณ์ในช่วงเวลาหนึ่งหรือช่วงเวลาระหว่างการเกิด

เหตุการณ์ และข้อมูลประเภทนี้มีลักษณะการกระจายข้อมูลเป็นการแจกแจงแบบแกมมา โดยที่ข้อมูลส่วนใหญ่กระจายตัวอยู่ใกล้ค่าศูนย์ ซึ่งแสดงถึงเหตุการณ์ที่ค่าที่มีค่าน้อยๆ หรือความผิดพลาดน้อยๆ จะเกิดขึ้นได้มาก เช่น ระยะเวลาที่ลูกค้ารอคอยจนกระทั่งได้รับบริการ อัตราเสี่ยงของการติดเชื้อโรค ปริมาณสารพิษในน้ำ เป็นต้น ผู้วิจัยจึงสนใจศึกษาข้อมูลที่มีการแจกแจงก่อนเป็นการแจกแจงแบบแกมมาซึ่งในวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์จะต้องใช้ทั้งความน่าจะเป็นก่อนและความน่าจะเป็นภายหลังเพื่อถ่วงน้ำหนักของแต่ละตัวแบบ ดังนั้นผู้วิจัยจึงใช้การแจกแจงก่อนและการแจกแจงความน่าจะเป็นภายหลังที่มีการแจกแจงแบบเดียวกัน เพื่อให้ทราบถึงประสิทธิภาพของวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์ ส่วนวิธีที่สามเป็นวิธีการพื้นฐานซึ่งนิยมใช้กันอย่างแพร่หลายและมีประสิทธิภาพดีในการคัดเลือกสมการถดถอยที่ดีที่สุดเมื่อการวิเคราะห์ความถดถอยเป็นไปตามข้อตกลงเบื้องต้น[4] ดังนั้นผู้วิจัยจึงสนใจศึกษาวิธีคัดเลือกตัวแบบถดถอยที่ดีที่สุดทั้ง 3 วิธี เพื่อศึกษาดูแนวโน้มว่าวิธีคัดเลือกตัวแบบถดถอยที่ดีที่สุดเชิงเบส์จะมีความเหมาะสมมากกว่าวิธีพื้นฐานหรือไม่และวิธีใดจะให้ค่าพยากรณ์ที่ถูกต้องและแม่นยำมากที่สุดเพื่อนำไปใช้ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณต่อไป

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบวิธีการคัดเลือกตัวแบบถดถอยที่ดีที่สุดในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณเมื่อใช้การแจกแจงก่อนแบบคู่สังยุคแกมมาจากทั้ง 3 วิธี ข้างต้นกล่าวคือ 1) วิธี BMA_{MC^3} 2) วิธี OPM 3) วิธี SR

1.3 ข้อตกลงเบื้องต้น

1. รูปแบบทั่วไปของสมการถดถอยเชิงเส้นพหุคูณมีรูปแบบดังสมการ (1.1)
2. ตัวแปรอิสระแต่ละตัวถือว่าเป็นค่าคงที่
3. ความคลาดเคลื่อนสุ่มเป็นตัวแปรสุ่มที่มีการแจกแจงเดียวกันและส่วนเบี่ยงเบนมาตรฐานเป็น σ เหมือนกันและเป็นอิสระซึ่งกันและกัน

1.4 สมมติฐานของการวิจัย

1. วิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณนั้น จะให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากกว่าวิธีการถดถอยแบบขั้นบันได

---

[4] จะเด็จ สวรรค์ตรานนท์, การเปรียบเทียบวิธีที่ใช้สำหรับการเลือกสมการถดถอยที่ดีที่สุด, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2530), หน้า 202.

2. วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ จะให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากกว่าวิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด

1.5 ขอบเขตของการวิจัย

1. ตัวแบบของการถดถอยเชิงเส้นพหุคูณที่สนใจศึกษาจะเป็นเชิงเส้นในพารามิเตอร์และเป็นเชิงเส้นในตัวแปรอิสระ

2. ตัวแบบการถดถอยในการวิจัยครั้งนี้เป็นตัวแบบติดกลุ่ม

3. ขนาดตัวอย่างที่ศึกษา คือ 15 30 50 และ 100 โดยที่ในกรณีขนาดตัวอย่างเท่ากับ 15 ผู้วิจัยได้ทำการศึกษาเฉพาะกรณีจำนวนตัวแปรอิสระเท่ากับ 3 และ 5 ส่วนในกรณีขนาดตัวอย่างเท่ากับ 30 ผู้วิจัยได้ทำการศึกษาเฉพาะกรณีจำนวนตัวแปรอิสระเท่ากับ 3 5 และ 10*

4. จำนวนตัวแปรอิสระที่ศึกษา คือ 3 5 10 และ 15

5. ค่าส่วนเบี่ยงเบนมาตรฐาน คือ 0.25 0.50 และ 2.50**

6. ในการวิจัยครั้งนี้จะศึกษากรณีการแจกแจงก่อนของเวกเตอร์พารามิเตอร์สัมประสิทธิ์ความถดถอย β แบบคู่สังยุคแกมมา

7. ค่าคงที่สำหรับวิธี BMA_{MC^3} และวิธี OPM กำหนดให้มี 4 ระดับ คือ (1,5) (1,10) (10,100) และ (10,500)***

8. ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติ-คาร์โลกระทำซ้ำ 500 รอบในแต่ละสถานการณ์

---

* ผู้วิจัยได้กำหนดขนาดตัวอย่างให้มากกว่าจำนวนตัวแปรอิสระอย่างน้อย 3 เท่า

** เนื่องจากค่าส่วนเบี่ยงเบนมาตรฐานจะมีผลต่อค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ดังนั้นเมื่อกำหนดค่าส่วนเบี่ยงเบนมาตรฐานที่มีค่าเปลี่ยนแปลงไม่มากจะทำให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่ได้มีค่าใกล้เคียงกัน ดังนั้นในกรณีที่กำหนดค่าส่วนเบี่ยงเบนมาตรฐานจาก 0.50 เป็น 2.50 นั้น ผู้วิจัยจึงได้กำหนดค่าส่วนเบี่ยงเบนมาตรฐานให้มีค่าเพิ่มขึ้นเพื่อแสดงให้เห็นถึงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าเพิ่มขึ้นตามค่าส่วนเบี่ยงเบนมาตรฐานที่เพิ่มขึ้น

*** จากงานวิจัยของจอร์จและแมคคัลลอชกล่าวว่าการกำหนดให้ค่า $\frac{\sigma_{\beta_i}}{\tau_i}$ และ c_i เป็นค่าคงที่ จะทำให้ได้ผลลัพท์ง่ายและรวดเร็วแล้ว การกำหนดค่า $(\frac{\sigma_{\beta_i}}{\tau_i}, c_i)$ เป็น 4 ระดับดังกล่าวจะแบ่งแยกการกระจายของพารามิเตอร์สัมประสิทธิ์การถดถอยได้อย่างชัดเจนซึ่งเมื่อกำหนดค่า $\frac{\sigma_{\beta_i}}{\tau_i}$ ให้มีค่ามากขึ้น การกระจายของพารามิเตอร์ก็จะมากขึ้น

1.6 คำจำกัดความต่าง ๆ ที่ใช้ในการวิจัย

1. ตัวแปรตาม (dependent variable (y)) หมายถึงตัวแปรที่เราสนใจที่เราต้องการศึกษา

2. ตัวแปรอิสระ (independent variable (x)) หมายถึงตัวแปรที่เราคิดว่ามีความสัมพันธ์กับตัวแปรตามที่เราต้องการศึกษา หรือมีผลกระทบต่อสิ่งที่เราต้องการศึกษา

3. ความคลาดเคลื่อนยกกำลังสอง (MSE) หมายถึงความคลาดเคลื่อนที่เกิดขึ้นจากการใช้สมการในการพยากรณ์สำหรับประมาณค่าตัวแปรตาม

4. ตัวแบบติดกลุ่ม (nested models) หมายถึงตัวแบบ 2 ตัวแบบจะติดกันถ้าในแต่ละพจน์ของตัวแบบแรกเป็นส่วนหนึ่งของตัวแบบที่สอง ซึ่งตัวแบบที่สองจะมีพจน์มากกว่าตัวแบบแรกอย่างน้อย 1 พจน์ ตัวแบบที่สองที่มีความซับซ้อนมากกว่าตัวแบบแรกจะเรียกว่าตัวแบบที่สมบูรณ์ (completed model) ส่วนตัวแบบแรกที่เป็นตัวแบบอย่างง่ายของตัวแบบสองเรียกว่าตัวแบบลดรูป (reduced model)

1.7 ขั้นตอนในการวิจัย

1. กำหนดส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนสุ่ม ขนาดตัวอย่าง จำนวนตัวแปรอิสระ ระดับนัยสำคัญและค่าคงที่สำหรับวิธี BMA_{MC^3} และวิธี OPM

2. สร้างข้อมูลตัวแปรตามจากตัวแปรอิสระและความคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

3. สร้างตัวแบบจากวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ 3 วิธี คือ

   1) วิธีการเฉลี่ยตัวแบบของเบส์ โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ (BMA_{MC^3})

   2) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (OPM)

   3) วิธีการถดถอยแบบขั้นบันได (SR)

4. หาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละตัวแบบ

5. สรุปผลในรูปของตารางและรูปภาพ

1.8 ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นแนวทางในการเลือกใช้วิธีการคัดเลือกตัวแบบที่ดีที่สุดภายใต้แนวทางของเบส์สำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณได้อย่างเหมาะสมเพื่อนำไปใช้ในการพยากรณ์ให้มีความถูกต้องและแม่นยำมากที่สุด

2. เป็นแนวทางสำหรับผู้สนใจที่จะศึกษาวิธีการคัดเลือกตัวแบบการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์สำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ

3. เป็นแนวทางในการศึกษาวิธีการคัดเลือกตัวแบบการถดถอยวิธีอื่นๆ ภายใต้แนวทางของเบส์

1.9 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบถดถอยวิธีใดนั้นจะมีประสิทธิภาพมากที่สุดพิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error (AMSE)) และการตัดสินใจจะใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error (RDAMSE)) เพื่อประกอบการเปรียบเทียบประสิทธิภาพของวิธีการของตัวสถิติต่างๆ มีสูตรดังนี้

$$MSE_j = \frac{\sum_{j=1}^{n}(y_i - \hat{y}_i)^2}{n-(p+1)}$$

$$\therefore \qquad AMSE = \frac{\sum_{j=1}^{500} MSE_j}{500}$$

เมื่อ y_i เป็นค่าสังเกตที่ i

$\hat{y}_i$ เป็นค่าพยากรณ์ที่ i

p เป็นจำนวนพารามิเตอร์ในตัวแบบถดถอย โดยยังไม่รวม β_0

n เป็นขนาดตัวอย่าง

MSE_j เป็นค่าคลาดเคลื่อนกำลังสองเฉลี่ยของการทำซ้ำรอบที่

และ $AMSE$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำ 500 รอบ

ส่วน $$RDAMSE_i = \frac{(AMSE_i - AMSE_{\min})}{AMSE_{\min}} \times 100\%$$

เมื่อ $AMSE_i$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากวิธีที่ i

และ $AMSE_{\min}$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าต่ำสุดจากทุกวิธี

อำนาจการพยากรณ์ของตัวแบบถดถอยจะลดลงถ้าค่าคลาดเคลื่อนกำลังสองเฉลี่ยเพิ่มขึ้น ดังนั้นในการเปรียบเทียบวิธีการคัดเลือกตัวแบบถดถอยที่ดีที่สุดจะพิจารณาว่าวิธีใดที่ให้ค่า

คลาดเคลื่อนกำลังสองต่ำที่สุดจะเป็นวิธีที่ดีที่สุด เนื่องจากค่าคลาดเคลื่อนกำลังสองเฉลี่ยใช้จำนวนพารามิเตอร์ในตัวแบบถดถอยเป็นตัวหาร ดังนั้นเมื่อตัวแบบที่ได้จากวิธีการเฉลี่ยตัวแบบจากทุกตัวแบบที่เป็นไปได้โดยใช้จำนวนพารามิเตอร์ที่เท่ากันกับวิธีการที่พิจารณาเพียงตัวแบบเดียวเป็นตัวหารในการคำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยอาจจะทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่ได้มีความแปรปรวนเกิดขึ้น