

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์ด้วยวิธีแบบเบสเชิงประจักษ์สำหรับตัวแบบCox' proportional hazardในกรณีที่ข้อมูลมีมิติสูง และทดสอบประสิทธิภาพของวิธีการดังกล่าวโดยสร้างกรณีศึกษาขึ้นมา แต่จะมีสถานการณ์ที่แตกต่างกัน โดยลักษณะข้อมูลที่สร้างขึ้นเป็นกรณีที่มักเกิดปัญหาในการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์ที่พบได้ทั่วไป เพื่อจะตรวจสอบว่าสถานการณ์ของข้อมูลลักษณะใดที่ส่งผลต่อประสิทธิภาพของวิธีแบบเบสเชิงประจักษ์ โดยในการศึกษาครั้งนี้จะทดสอบเฉพาะเมื่อกำหนดค่าสัมประสิทธิ์เริ่มต้นด้วยค่าจริงและค่าประมาณที่ได้จากวิธี Lasso เท่านั้น พร้อมทั้งเปรียบเทียบประสิทธิภาพกับวิธี Lasso เมื่อกำหนดให้ทดสอบในสถานการณ์ที่เหมือนกันและข้อมูลชุดเดียวกัน ผลการทดลองมีข้อสรุปดังนี้

5.1. สรุปผลการทดลอง

จากการศึกษาเพื่อทดสอบประสิทธิภาพการทำงานของวิธีแบบเบสเชิงประจักษ์ โดยแบ่งการทดสอบเป็น 2 ส่วนคือ

5.1.1 ทดสอบหาปัจจัยที่ส่งผลต่อประสิทธิภาพในการทำงานของวิธีแบบเบสเชิงประจักษ์

5.1.2 เปรียบเทียบประสิทธิภาพของวิธีแบบเบสเชิงประจักษ์กับวิธี Lasso โดยผลการทดสอบทั้งสองกรณีมีรายละเอียดดังนี้

5.1.1 ทดสอบหาปัจจัยที่ส่งผลต่อประสิทธิภาพในการทำงานของวิธีแบบเบสเชิงประจักษ์

ในการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์มักเกิดปัญหาเมื่อข้อมูลที่พิจารณา มีสถานการณ์ที่ไม่เอื้อต่อการทำงานของวิธีที่ใช้ ในการศึกษาครั้งนี้ได้แยกสถานการณ์ที่คาดว่าจะส่งผลกระทบต่อวิธีการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์ด้วยวิธีแบบเบสเอาไว้ 3 กรณี คือ กรณีที่ข้อมูลถูกเซ็นเซอร์ที่ระดับต่างๆ, กรณีที่จำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่างโดยแบ่งเป็น 3 ระดับคือจำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่างเล็กน้อย (100:300) มากกว่า

ปานกลาง (100:500) และมากกว่ามาก (100:1000) และผลกระทบสุดท้ายที่คาดว่าจะส่งผลต่อการทำงานคือ การกำหนดค่าสัมประสิทธิ์เริ่มต้น จากการทดลองสามารถสรุปผลได้ดังนี้

ความผิดพลาดในการตรวจจับเชิงบวก

สิ่งที่ส่งผลต่ออัตราความผิดพลาดในการตรวจจับเชิงบวกคือ ระดับข้อมูลเซ็นเซอร์ ดังที่ปรากฏในผลการทดลอง ที่ร้อยละของข้อมูลเซ็นเซอร์ต่ำ (10%) จะให้ค่าความผิดพลาดเชิงบวกต่ำกว่ากรณีร้อยละของข้อมูลเซ็นเซอร์กลาง (50%) และร้อยละของข้อมูลเซ็นเซอร์สูง (70%) ตามลำดับ กล่าวคือ เมื่อจำนวนข้อมูลที่ถูกเซ็นเซอร์เพิ่มมากขึ้น อัตราความผิดพลาดในการตรวจจับเชิงบวกก็จะสูงขึ้นด้วย ปัจจัยต่อมาคือ อัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระคือ เมื่อพิจารณาที่ขนาดตัวอย่างเท่ากัน ถ้าตัวแปรอิสระเพิ่มมากขึ้น อัตราความผิดพลาดในการตรวจจับเชิงบวกก็จะมีแนวโน้มสูงขึ้น และปัจจัยสุดท้ายคือ การกำหนดค่าสัมประสิทธิ์เริ่มต้น โดยเมื่อกำหนดค่าสัมประสิทธิ์เริ่มต้นเป็นค่าจริง จะทำให้วิธีแบบเบสเชิงประจักษ์มีประสิทธิภาพมากกว่าการกำหนดค่าสัมประสิทธิ์เริ่มต้นด้วยค่าประมาณที่ได้จากวิธี Lasso

ความผิดพลาดในการตรวจจับเชิงลบ

สิ่งที่ส่งผลต่ออัตราความผิดพลาดในการตรวจจับเชิงลบคือ ระดับข้อมูลเซ็นเซอร์ ดังที่ปรากฏในผลการทดลอง ที่ร้อยละของข้อมูลเซ็นเซอร์ต่ำ (10%) จะให้ค่าความผิดพลาดเชิงลบต่ำกว่ากรณีร้อยละของข้อมูลเซ็นเซอร์กลาง (50%) และร้อยละของข้อมูลเซ็นเซอร์สูง (70%) ตามลำดับ กล่าวคือ เมื่อจำนวนข้อมูลที่ถูกเซ็นเซอร์เพิ่มมากขึ้น อัตราความผิดพลาดในการตรวจจับเชิงลบก็จะสูงขึ้นด้วย ปัจจัยต่อมาคือ อัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระคือ ยิ่งอัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระต่ำ (100:1000) อัตราความผิดพลาดในการตรวจจับเชิงลบจะสูงตาม และปัจจัยสุดท้ายคือ การกำหนดค่าสัมประสิทธิ์เริ่มต้น โดยเมื่อกำหนดค่าสัมประสิทธิ์เริ่มต้นเป็นค่าจริง จะทำให้วิธีแบบเบสเชิงประจักษ์มีประสิทธิภาพมากกว่าการกำหนดค่าสัมประสิทธิ์เริ่มต้นด้วยค่าประมาณที่ได้จากวิธี Lasso

จากผลการทดลองที่เกิดขึ้น ยังสามารถสรุปได้ว่าสิ่งที่ส่งผลต่ออัตราความผิดพลาดในการตรวจจับเชิงลบของการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์ด้วยวิธีแบบเบสเชิงประจักษ์มากที่สุดคือ การกำหนดค่าสัมประสิทธิ์เริ่มต้น เนื่องจากถ้ากำหนดให้ค่าสัมประสิทธิ์เริ่มต้นคือค่าจริงแล้วไม่ว่าอัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระหรือข้อมูลเซ็นเซอร์จะเป็นเท่าไร

อัตราความผิดพลาดในการตรวจจับเชิงลบจะเท่ากับศูนย์เสมอ ส่วนอีกสองปัจจัยแม้จะส่งผลต่อความผิดพลาดในการตรวจจับเชิงลบ แต่ผลก็ไม่เห็นชัดเท่าปัจจัยการกำหนดค่าเริ่มต้น

สรุป

จากค่าทดสอบประสิทธิภาพวิธีแบบเบสเชิงประจักษ์ ทั้งสองคืออัตราความผิดพลาดในการตรวจจับเชิงบวกและอัตราในการตรวจจับเชิงลบ ต่างให้แนวโน้มของผลการทดสอบเหมือนกัน คือ ทั้ง 3 ปัจจัย ได้แก่ ร้อยละของข้อมูลเซ็นเซอร์, อัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระ และการกำหนดสัมประสิทธิ์เริ่มต้น ต่างส่งผลต่อประสิทธิภาพการทำงานของวิธีแบบเบสเชิงประจักษ์ ซึ่งเมื่อเปรียบเทียบผลที่เกิดขึ้นจะเห็นว่าทั้ง 3 ปัจจัย ส่งผลกระทบต่ออัตราความผิดพลาดในการตรวจจับเชิงบวก สูงกว่าอัตราในการตรวจจับเชิงลบ โดยเฉพาะอย่างยิ่งร้อยละของข้อมูลเซ็นเซอร์ที่ยังมีค่าสูงเท่าไร อัตราความผิดพลาดก็จะเพิ่มความแตกต่างจากเดิมอย่างเห็นได้ชัด

5.1.2 เปรียบเทียบประสิทธิภาพของวิธีแบบเบสเชิงประจักษ์กับวิธี Lasso

ผลการเปรียบเทียบวิธีคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์สำหรับตัวแบบ Cox's proportional hazard ที่ข้อมูลมีมิติสูงระหว่างวิธีแบบเบสเชิงประจักษ์และวิธี Lasso พบว่าเมื่อพิจารณาอัตราความผิดพลาดในการตรวจจับเชิงลบของทั้ง 2 วิธี มีผลใกล้เคียงกัน โดยภาพรวมถือว่าวิธีแบบเบสเชิงประจักษ์ให้อัตราความผิดพลาดเชิงลบต่ำกว่า สำหรับอัตราความผิดพลาดในการตรวจจับเชิงบวก ผลที่ได้จะมีความแตกต่างกันมาก โดยวิธีแบบเบสเชิงประจักษ์จะมีอัตราความผิดพลาดในการตรวจจับเชิงลบต่ำกว่าวิธี Lasso แต่ในกรณีศึกษาที่ 8 เมื่ออัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระ เป็น 100:1000 ข้อมูลเซ็นเซอร์ 50% และกรณีที่ 9 เมื่ออัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระ เป็น 100:1000 ข้อมูลเซ็นเซอร์ 70% อัตราความผิดพลาดในการตรวจจับเพิ่มสูงขึ้นมาก ดังนั้น มีแนวโน้มว่า เมื่อจำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่างมากขึ้น และร้อยละข้อมูลเซ็นเซอร์เพิ่มขึ้น มีแนวโน้มว่าจะส่งผลต่อประสิทธิภาพของวิธีแบบเบส

5.2. ข้อเสนอแนะ

ผลการวิจัยครั้งนี้มีข้อเสนอแนะแบ่งเป็น 2 ด้าน คือ

5.2.1 ข้อเสนอแนะด้านการนำไปใช้งาน

จากผลการศึกษาพบว่า การคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์ด้วยวิธีแบบเบสเชิงประจักษ์สำหรับตัวแบบ Cox's proportional hazard มีประสิทธิภาพกว่าวิธี Lasso ทั้งในด้านของความถูกต้อง แม่นยำ แม้ในสถานการณ์ที่มีจำนวนตัวแปรอิสระมากและระดับของร้อยละของข้อมูลเซ็นเซอร์สูง และในด้านระยะเวลาของกระบวนการทำงาน พบว่าวิธีแบบเบสเชิงประจักษ์ยังใช้เวลาในการทำงานน้อยกว่าวิธี Lasso มาก ดังนั้น วิธีแบบเบสเชิงประจักษ์ถือเป็นทางเลือกหนึ่งที่เหมาะสมสำหรับการคัดเลือกตัวแปรและประมาณค่าสัมประสิทธิ์ในสถานการณ์ที่ข้อมูลบางส่วนเป็นข้อมูลเซ็นเซอร์ หรือเมื่อจำนวนตัวแปรอิสระมีมากกว่าขนาดตัวอย่าง (ข้อมูลมิติสูง)

5.2.1 ข้อเสนอแนะด้านการศึกษาวิจัย

เพื่อเป็นแนวทางให้ผู้สนใจได้ศึกษาเพิ่มเติม ในการศึกษาครั้งต่อไป อาจทำการศึกษาในกรณีต่างๆ ดังนี้

1. เนื่องจากการทดสอบการศึกษาในครั้งนี้ กำหนดให้ตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน ในความเป็นจริง เหตุการณ์เช่นนี้มักเกิดได้น้อย หรือแทบไม่เกิดขึ้นเลย ดังนั้นในการศึกษาครั้งต่อไป อาจเพิ่มให้ตัวแปรอิสระมีความสัมพันธ์กันในลักษณะต่างๆ
2. เพื่อประโยชน์ของการนำไปใช้งาน อาจเพิ่มขนาดตัวอย่างต่อตัวแปรอิสระให้มีความแตกต่างกันมากขึ้น เนื่องจากการศึกษาครั้งนี้ อาจพิจารณาปัจจัยระหว่างอัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระที่ระดับไม่แตกต่างกันมากนัก จึงไม่สามารถบอกขีดจำกัดของวิธีแบบเบสเชิงประจักษ์ว่าจะมีประสิทธิภาพมากกว่าวิธี Lasso ไปจนถึงอัตราส่วนระหว่างขนาดตัวอย่างต่อตัวแปรอิสระเป็นเท่าไร
3. ทดสอบวิธีแบบเบสเชิงประจักษ์กรณีที่ข้อมูลมิติสูงกับตัวแบบอื่นๆ