

บทที่ 1 บทนำ

1.1 กล่าวนำ

ในการประยุกต์ใช้สถิติกับงานทดลองวิจัยด้านต่างๆ ไม่ว่าจะเป็นด้านพืช สัตว์ เศรษฐศาสตร์ ตลอดจนงานทดลองวิจัยเกี่ยวกับระบาดวิทยา (epidemiology) มักจะพบอยู่บ่อยๆ ว่า ผลจากการทดลองที่ได้ จะให้ค่าเป็นหนึ่งในสองอย่างที่เป็นไปได้ทั้งหมด เช่น ในการทดลองเพาะเมล็ดพันธุ์พืช จะได้เมล็ดงอกหรือไม่งอก ในการทดลองรักษาสัตว์ที่เป็นโรค จะได้ผลว่าสัตว์หายหรือไม่หายจากโรค ในการทดลองใช้ยาฆ่าแมลง จะได้ว่าแมลงตายหรือไม่ตาย เป็นต้น ข้อมูลที่ได้จากการทดลองที่มีลักษณะดังกล่าวนี้เรียกว่า ข้อมูลทวิ (binary data or quantal) และจะนิยามค่าที่ได้แต่ละค่าว่า เกิดผลสำเร็จ (success) กับไม่เกิดผลสำเร็จ (failure)

การวิเคราะห์การถดถอยโลจิสติก (logistic regression analysis) เป็นการวิเคราะห์การถดถอยสำหรับข้อมูลทวิ ที่ใช้โมเดลโลจิสติก (logistic model) เป็นโมเดลในการทำนายค่าของตัวแปรตาม ซึ่งอยู่ในรูปของความน่าจะเป็น และแม้ว่าจะมีโมเดลอื่น ๆ ที่สามารถนำมาใช้เป็นโมเดลการถดถอย ในการวิเคราะห์การถดถอยสำหรับข้อมูลทวิได้ก็ตาม เช่น โมเดลเชิงเส้น (linear model) โมเดลโพรบิท (probit model) และโมเดล complementary log-log เป็นต้น แต่ก็พบว่านิยมนำใช้โมเดลโลจิสติกกันมากกว่าโมเดลอื่น ๆ ทั้งนี้เนื่องจากการใช้โมเดลโลจิสติกเป็นโมเดลทำนายค่า สามารถนำไปประยุกต์ใช้ได้หลากหลายรูปแบบกว่า และในการประมาณค่าพารามิเตอร์ของโมเดลโดยวิธีภาวะน่าจะเป็นสูงสุด จะให้ตัวประมาณที่มีการแจกแจงที่ยอมรับได้ อันเป็นประโยชน์ต่อการอนุมานทางสถิติ

แม้ว่า การใช้โมเดลโลจิสติกเป็นโมเดลในการวิเคราะห์การถดถอย จะมีความเหมาะสมในหลายประการก็ตาม แต่มีอุปสรรคหนึ่งเกี่ยวกับการวิเคราะห์ประมวลผล คือ เป็นเรื่องยากหรือแทบเป็นไปไม่ได้เลยที่จะวิเคราะห์ประมวลผลด้วยมือ อย่างไรก็ตาม ในปัจจุบัน

ปัญหานี้มิใช่อุปสรรคอีกต่อไปแล้ว เนื่องจากมีเครื่องคอมพิวเตอร์และโปรแกรมสำเร็จรูปทางสถิติช่วยในการวิเคราะห์ประมวลผลได้เป็นอย่างดี

1.2 ประวัติความเป็นมา

โมเดลการถดถอยโลจิสติก ถูกแนะนำเป็นครั้งแรกโดย Berkson (1944) ซึ่งแสดงให้เห็นว่า สามารถวิเคราะห์โมเดลด้วยวิธีกำลังสองต่ำสุดแบบถ่วงน้ำหนักที่ทำซ้ำ (iteratively weighted least squares) ได้อย่างไร Changers and Cox (1967) ได้ทำการเปรียบเทียบระหว่างการแปลงเป็น logistic กับการแปลงเป็น probit และ Berkson (1951) อธิบายว่า ทำไมเขาจึงชอบ logistic มากกว่า probit

Thisted (1988) อธิบายให้เห็นถึงความสอดคล้องกันระหว่างวิธีกำลังสองต่ำสุดแบบถ่วงน้ำหนักที่ทำซ้ำ กับวิธีภาวะน่าจะเป็นสูงสุด ซึ่งใช้กับโมเดลการถดถอยโลจิสติก และ Everitt (1987) แสดงวิธีการประมาณค่าพารามิเตอร์แบบภาวะน่าจะเป็นสูงสุด ด้วยวิธี scoring ของ Fisher (1925) Nelder and Wedderburn (1972) แสดงวิธีการใช้ algorithm เพื่อให้ได้ค่าประมาณแบบภาวะน่าจะเป็นสูงสุดอยู่ในรูปแบบทั่ว ๆ ไป และได้พัฒนาขึ้นใหม่โดย McCullagh and Nelder (1989) และ Aitken et al (1989)

การใช้ deviance เพื่อตรวจสอบภาวะสารูปสันติ ถูกนำเสนอเป็นครั้งแรกโดย Nelder and Wedderburn (1972) McCullagh and Nelder (1989) เสนอรายละเอียดเกี่ยวกับความเหมาะสมของการประมาณการแจกแจงของสมมติฐานัล ด้วยการแจกแจงแบบไคสแควร์ และ Collett (1991) รวบรวมและให้รายละเอียดเกี่ยวกับการใช้โมเดลสำหรับการวิเคราะห์ข้อมูลทวิ

1.3 ขอบเขตของการศึกษา

ในการค้นคว้าอิสระครั้งนี้ จะเป็นการแนะนำให้รู้จักกับการใช้โมเดลโลจิสติกในการ

วิเคราะห์การถดถอยสำหรับข้อมูลทวิ ที่มาจากการแจกแจงแบบทวินาม แบบทั่ว ๆ ไป ซึ่งจะได้มีการทบทวนการวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลแบบต่อเนื่อง ในบทที่ 2 แล้วประยุกต์ใช้การวิเคราะห์การถดถอยเชิงเส้นกับข้อมูลทวิ ในบทที่ 3 หลังจากนั้นในบทที่ 4 จะเป็นการอธิบายถึงการใช้โมเดลโลจิสติก ในการวิเคราะห์การถดถอยสำหรับข้อมูลทวิที่มาจากการแจกแจงแบบทวินาม ซึ่งจะมีรายละเอียดเกี่ยวกับการประมาณค่าพารามิเตอร์ โดยวิธีภาวะน่าจะเป็นสูงสุด และวิธีกำลังสองต่ำสุดแบบถ่วงน้ำหนักที่ซ้ำ และในบทที่ 5 จะแสดงการวิเคราะห์โลจิสติกจากข้อมูลตัวอย่าง ด้วยโปรแกรมสำเร็จรูป SAS