

ทัศนีย์ น้าเจริญ 2557: การวิเคราะห์การจำแนกข้อมูลที่ไม่สมดุลในแต่ละกลุ่ม
กรณีศึกษากลุ่มเสี่ยงโรคเบาหวาน ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ) สาขาสถิติ
ภาควิชาสถิติ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์ลลิต อิงศรีสว่าง,
Ph.D. 99 หน้า

การศึกษานี้เป็นการเปรียบเทียบประสิทธิภาพของวิธีการวิเคราะห์การจำแนกที่อิงพารามิเตอร์และการวิเคราะห์การจำแนกที่ไม่อิงพารามิเตอร์ ในการจำแนกข้อมูล 3 กลุ่ม โดยที่มีความไม่สมดุลของข้อมูลแต่ละกลุ่ม ข้อมูลที่นำมาศึกษาเป็นข้อมูลกลุ่มเสี่ยงโรคเบาหวานของเจ้าหน้าที่ที่ปฏิบัติงานในโรงพยาบาลรัฐแห่งหนึ่งจำนวน 599 คน มีการวินิจฉัยเป็น 3 กลุ่ม คือ กลุ่มไม่เสี่ยงโรคเบาหวาน มีประมาณ 90% กลุ่มเสี่ยงโรคเบาหวาน มีประมาณ 5% และกลุ่มเป็นโรคเบาหวาน มีประมาณ 5% และมีตัวแปรอิสระที่นำมาพิจารณาประกอบด้วย อายุ เพศ, ระดับน้ำตาลในเลือดและดัชนีมวลกาย ซึ่งข้อมูลดังกล่าวไม่มีการแจกแจงแบบปกติพหุและความแปรปรวนร่วมแต่ละกลุ่มไม่เท่ากัน วิธีการศึกษาทำการบดสเตรปข้อมูลกลุ่มเสี่ยงโรคเบาหวานจำนวน 50 ชุด ชุดละ 599 ค่าสังเกต แต่ละชุดจะทำการวิเคราะห์การจำแนกที่อิงพารามิเตอร์ด้วยฟังก์ชันการจำแนกเชิงเส้นและฟังก์ชันการจำแนกกำลังสอง และวิเคราะห์การจำแนกที่ไม่อิงพารามิเตอร์ด้วยวิธีเคเนียร์สเนเบอร์ เมื่อกำหนด k เป็น 3, 4, 5 พร้อมทั้งกำหนดค่าความน่าจะเป็นก่อนหน้าด้วยสัดส่วนที่เท่ากันและสัดส่วนที่แตกต่างกันสำหรับสร้างกฎการจำแนกกลุ่มที่เหมาะสม และวัดประสิทธิภาพของวิธีการวิเคราะห์การจำแนกแต่ละวิธีด้วยค่าเฉลี่ยอัตราความผิดพลาด ผลการศึกษาพบว่าจากข้อมูลบดสเตรป 50 ชุด การวิเคราะห์การจำแนกด้วยวิธีเคเนียร์สเนเบอร์ เมื่อ $k=3$ หรือ 4 และเมื่อกำหนดค่าความน่าจะเป็นก่อนหน้าของกลุ่มไม่เสี่ยงโรคเบาหวาน: กลุ่มเสี่ยงโรคเบาหวาน : กลุ่มเป็นโรคเบาหวาน ด้วยสัดส่วนที่เหมือนกับข้อมูลจริงคือ 0.90 : 0.05 : 0.05 หรือ 0.80 : 0.10 : 0.10 จะได้ค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกกลุ่มผิดมีค่าต่ำที่สุด

ลายมือชื่อนิสิต

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก