

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม 3 วิธี คือ วิธีการจำแนกกลุ่มแบบดั้งเดิม วิธีเค-เน็ยเรสเนเบอร์ และวิธีวิเคราะห์องค์ประกอบหลักแบบเคอร์เนล กรณีข้อมูลมีการแจกแจงแบบปกติเชิงพหุและลอกนอร์มอลเชิงพหุ โดยกำหนดขนาดตัวอย่างที่ใช้ในการศึกษาเป็น 100, 200 และ 500 จำนวนตัวแปรอิสระเป็น 4, 6 และ 8 ระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่ใช้แบ่งกลุ่มเป็น 3 ระดับ คือ ระดับต่ำ ปานกลาง และสูง และอัตราส่วนของข้อมูลระหว่าง Training: Validation เป็น 90:10, 80:20, 70:30, 60:40 และ 50:50 ทำการจำลองข้อมูลโดยกระทำซ้ำ 5,000 ครั้ง ในแต่ละสถานการณ์ที่กำหนด และใช้อัตราความผิดพลาดของการจำแนกเป็นเกณฑ์ในการเปรียบเทียบ

ผลการวิจัยพบว่า เมื่อขนาดตัวอย่างและสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้น อัตราการจำแนกกลุ่มข้อมูลผิดพลาดของทั้ง 3 วิธี มีแนวโน้มลดลง โดยกรณีที่ข้อมูลมีการแจกแจงแบบปกติเชิงพหุ พบว่าเมื่อสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับต่ำ วิธีการจำแนกกลุ่มแบบดั้งเดิมให้อัตราการจำแนกกลุ่มข้อมูลผิดพลาดต่ำที่สุด และเมื่อสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับปานกลางและสูง พบว่า วิธีเค-เน็ยเรสเนเบอร์และวิธีวิเคราะห์องค์ประกอบหลักแบบเคอร์เนลให้อัตราการจำแนกกลุ่มข้อมูลผิดพลาดใกล้เคียงกันและมีค่าต่ำที่สุดเป็นส่วนใหญ่ และกรณีที่ข้อมูลมีการแจกแจงแบบลอกนอร์มอลเชิงพหุ พบว่าเมื่อสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับต่ำ ปานกลางและสูง โดยส่วนใหญ่แล้วพบว่าวิธีเค-เน็ยเรสเนเบอร์ให้อัตราการจำแนกกลุ่มข้อมูลผิดพลาดต่ำที่สุด โดยเมื่อสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับปานกลางและสูง วิธีเค-เน็ยเรสเนเบอร์และวิธีวิเคราะห์องค์ประกอบหลักแบบเคอร์เนลให้อัตราการจำแนกกลุ่มข้อมูลผิดพลาดใกล้เคียงกัน

The objective of this study is to compare the efficiency of three discriminant methods: classical discriminant method, K-nearest neighbor method (KNN) and kernel principal components analysis method (KPCA) when data are distributed as multivariate normal and lognormal with the sample sizes 100, 200 and 500. The number of independent variables considered here consists of 4, 6 and 8, each with three levels of correlation coefficients: low, moderate and high. The ratios of Training : Validation used in this study are 90:10, 80:20, 70:30, 60:40 and 50:50. In each situation, the data are simulated and repeated 5,000 times. Rate of misclassification is used as a criterion for comparison.

The results indicate that rates of misclassification tend to be lowered as the sample sizes and correlation coefficients increase. With multivariate normal data, the classical discriminant method produces the smallest misclassification rate at low level of correlation coefficient, while KNN and KPCA mostly produce similar results and tend to yield the smallest misclassification rate at moderate and high levels of correlation coefficient. With multivariate lognormal distribution, KNN tends to produce the smallest misclassification rate for all three levels of correlation coefficient. In addition, the misclassification rates of KPCA are mostly close to those of KNN, especially at moderate and high levels of correlation coefficient.