

อาทิที่ ลาเต๊ะ : การเปรียบเทียบวิธีการทางสถิติและวิธีการโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกกรณี 2 กลุ่ม (A COMPARISON OF STATISTICAL AND LINEAR PROGRAMMING APPROACHES FOR TWO-GROUP DISCRIMINANT ANALYSIS) อาจารย์ผู้ควบคุมวิทยานิพนธ์ : รศ.ดร.สุดา ตรีการเฉลิมศักดิ์ และ ผศ.ดร.ปราณี นิลกรณ์. 144 หน้า. ISBN 974-464-638-1.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบอัตราการจัดแนกผิดในการวิเคราะห์การจำแนกกรณี 2 กลุ่มระหว่างฟังก์ชันจำแนกเชิงเส้นของฟิชเชอร์ (FLDF: Fisher 1936) กับตัวแบบการโปรแกรมเชิงเส้น 3 ตัวแบบ คือ ตัวแบบ MSD-FG2 (Freed and Glover 1986a) ตัวแบบ LCM1 (Lam et al. 1996) และตัวแบบ Extended DEA-DA (Sueyoshi 2001) โดยข้อมูลที่ศึกษามี 2 ลักษณะ คือ 1) ข้อมูลจริงซึ่งเป็นข้อมูลทฤษฎีเกี่ยวกับอัตราส่วนทางการเงินของบริษัทธุรกิจที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย 2) ข้อมูลจากการจำลองแบบจากการแจกแจงแบบปกติของตัวแปรพหุนา 3 ตัวแปรซึ่งแบ่งเป็น 2 กรณี คือ กรณีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรอิสระของทั้งสองกลุ่มเท่ากัน และกรณีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรอิสระของทั้งสองกลุ่มไม่เท่ากัน โดยในแต่ละกรณีประกอบด้วยปัจจัยต่อไปนี้ 2.1) กรณีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรอิสระของทั้งสองกลุ่มเท่ากันค่าความแปรปรวนของตัวแปรอิสระของค่าสังเกตปกติพิจารณาที่ 3 ระดับคือ 0.8, 1.0 และ 2.0 ส่วนกรณีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรอิสระของทั้งสองกลุ่มไม่เท่ากัน ค่าความแปรปรวนของตัวแปรอิสระของค่าสังเกตปกติของกลุ่มที่ 1 ต่อกลุ่มที่ 2 พิจารณาที่ 3 ระดับคือ 1:2, 1:4 และ 1:6 2.2) ความแปรปรวนร่วมระหว่างตัวแปรอิสระของค่าสังเกตปกติ 3 ระดับคือ 0.2, 0.5 และ 0.8 2.3) ความแปรปรวนของตัวแปรอิสระของค่าสังเกตปกติ 3 ระดับคือ 2, 3 และ 4 โดยมีความแปรปรวนร่วมระหว่างตัวแปรอิสระเท่ากับ 0.5 ในทุกระดับของความแปรปรวนของค่าสังเกตปกติ 2.4) กรณีไม่มีการปนเปื้อนค่าสังเกตปกติและกรณีปนเปื้อนค่าสังเกตปกติ 3 ระดับคือ 5%, 10% และ 15% ในแต่ละกรณีทำซ้ำ 100 ครั้ง โดยข้อมูลจริงประเมินจากจำนวนบริษัทที่จำแนกผิดจากทั้งหมด 210 บริษัท ส่วนข้อมูลจำลองแบบประเมินจากค่าเฉลี่ยของอัตราการจัดแนกผิดใน holdout sample ขนาด 200 ค่าสังเกต (100 ค่าสังเกตจากแต่ละกลุ่ม)

ผลการวิจัยพบว่ากรณีข้อมูลจริง ตัวแบบ LCM1 ให้อัตราการจัดแนกผิดต่ำสุด รองลงมาคือตัวแบบ FLDF ตัวแบบ Extended DEA-DA และตัวแบบ MSD-FG2 ตามลำดับ กรณีข้อมูลจำลองแบบ พบว่าเมื่อความแปรปรวน ความแปรปรวนร่วมของค่าสังเกตปกติและเปอร์เซ็นต์ของค่าสังเกตปกติเพิ่มขึ้น ทุกตัวแบบจะให้ค่าเฉลี่ยของอัตราการจัดแนกผิดเพิ่มขึ้น เมื่อความแปรปรวนของค่าสังเกตปกติเพิ่มขึ้น ทุกตัวแบบให้ค่าเฉลี่ยของอัตราการจัดแนกผิดลดลง กรณีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรอิสระของทั้งสองกลุ่มเท่ากัน ตัวแบบ MSD-FG2 ให้ค่าเฉลี่ยของอัตราการจัดแนกผิดต่ำสุด รองลงมาคือตัวแบบ Extended DEA-DA ตัวแบบ FLDF และตัวแบบ LCM1 ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ยของอัตราการจัดแนกผิดของทั้ง 4 วิธีตามระดับการปนเปื้อนค่าสังเกตปกติ พบว่ากรณีไม่มีการปนเปื้อนค่าสังเกตปกติและปนเปื้อนค่าสังเกตปกติ 5% ค่าเฉลี่ยของอัตราการจัดแนกผิดของทั้ง 4 วิธีไม่มีความแตกต่างกันอย่างมีนัยสำคัญ เมื่อเปอร์เซ็นต์ค่าสังเกตปกติเพิ่มขึ้นเป็น 10% และ 15% ตัวแบบ MSD-FG2 ให้ผลดีกว่าตัวแบบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ตัวแบบ Extended DEA-DA กับตัวแบบ FLDF ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ตัวแบบ Extended DEA-DA ให้ผลดีกว่าตัวแบบ LCM1 ส่วนตัวแบบ FLDF มีแนวโน้มที่จะให้ผลดีกว่าตัวแบบ LCM1 เมื่อเปอร์เซ็นต์ค่าสังเกตปกติมีค่าเพิ่มขึ้น กรณีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรอิสระของทั้งสองกลุ่มไม่เท่ากัน ตัวแบบ MSD-FG2 ให้ค่าเฉลี่ยของอัตราการจัดแนกผิดต่ำสุด รองลงมาคือตัวแบบ FLDF ตัวแบบ LCM1 และตัวแบบ Extended DEA-DA ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ยของอัตราการจัดแนกผิดของทั้ง 4 วิธีตามระดับการปนเปื้อนค่าสังเกตปกติ พบว่ากรณีไม่มีการปนเปื้อนค่าสังเกตปกติและปนเปื้อนค่าสังเกตปกติ 5% ตัวแบบ FLDF ให้ผลดีกว่าตัวแบบ MSD-FG2 อย่างมีนัยสำคัญ แต่เมื่อเปอร์เซ็นต์ค่าสังเกตปกติเพิ่มขึ้นเป็น 10% และ 15% ตัวแบบ MSD-FG2 ให้ผลดีกว่าตัวแบบ FLDF อย่างมีนัยสำคัญ อย่างไรก็ตามเมื่อเปรียบเทียบระหว่างตัวแบบ Extended DEA-DA กับตัวแบบ LCM1 พบว่าตัวแบบ Extended DEA-DA มีแนวโน้มที่จะให้ผลดีกว่าตัวแบบ LCM1 เมื่อเปอร์เซ็นต์ค่าสังเกตปกติเพิ่มขึ้น

AFIFI LATEH: A COMPARISON OF STATISTICAL AND LINEAR PROGRAMMING APPROACHES FOR TWO-GROUP DISCRIMINANT ANALYSIS. THESIS ADVISORS: ASSO.PROF. SUDA TRAGANTALERNGSAK, Ph.D., AND ASST.PROF. PRANEE NILAKORN, Ph.D. 144 pp. ISBN 974-464-638-1.

The objective of this research is to compare the misclassification rate of Fisher's Linear Discriminant Function (FLDF: Fisher 1936) and those of the three linear programming models (MSD-FG2: Freed and Glover 1986a, LCM1: Lam *et al.* 1996 and Extended DEA-DA: Sueyoshi 2001) for two-group discriminant analysis. The data sets under study included both real and simulated data. Real data were financial ratios of 210 Thai listed companies, 30 bankrupts and 180 nonbankrupts and the performance of each model was assessed by misclassification rate. Simulated data sets were generated from 3-variate normal distribution with equal and unequal variance-covariance matrices under the following conditions: i) for the equal variance-covariance matrix case, common variances of independent variables for clean observations were varied at 3 values, i.e. 0.8, 1.0 and 2.0 and for the unequal variance-covariance matrix case, the ratios of variances between two groups were set to 1:2, 1:4 and 1:6; ii) covariances between independent variables for clean observations were set to 0.2, 0.5, and 0.8; iii) variances of independent variables for outliers were set to 2, 3 and 4 with 0.5 covariance; iv) the percentage contamination levels were 0, 5, 10, and 15. The misclassification rate was calculated from 200 observations in holdout sample, 100 observations for each group. For each condition, 100 replications were performed to obtain the average misclassification rate for each model.

The findings were as follow: 1) For real data: the LCM1 model performed best, followed by FLDF, Extended DEA-DA and MSD-FG2 models, respectively. 2) For simulated data, in the equal variance-covariance matrix case, the MSD-FG2 model performed best, followed by Extended DEA-DA, FDLF and LCM1 models, respectively. All four models performed similarly at the 0% and 5% contamination levels. At the 10% and 15% contamination levels, the MSD-FG2 model significantly outperformed than the other three models, the Extended DEA-DA and FDLF models performed similarly and the Extended DEA-DA significantly outperformed than LCM1 model. FDLF model tended to perform better than LCM1 model as the level of contamination got higher. In the unequal variance-covariance matrix case, the MSD-FG2 model performed best, followed by FLDF, LCM1 and Extended DEA-DA models, respectively. At the 0% and 5% contamination levels, the FLDF model significantly outperformed than MSD-FG2 model, but at 10% and 15%, the MSD-FG2 model performed better than FLDF model. However, Extended DEA-DA model tended to perform better than LCM1 model as the level of contamination got higher.