

ศยามล ลำลองรัตน์ 2550: การศึกษาเปรียบเทียบการวินิจฉัยโรคตับด้วยการวิเคราะห์จำแนกกลุ่ม และการวิเคราะห์ถดถอยโลจิสติก กรณีศึกษาผู้ป่วยโรงพยาบาลเมืองฉะเชิงเทรา ปรินญาวิทยาศาสตร์ มหาวิทยาลัย (สถิติ) สาขาสถิติ ภาควิชาสถิติ ประชานกรรมการที่ปรึกษา: ผู้ช่วยศาสตราจารย์ บุญอ้อม โฉมที, Ph.D. 91 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการวินิจฉัยการเป็น หรือ ไม่เป็น โรคตับของผู้ป่วยโรงพยาบาลเมืองฉะเชิงเทราที่ได้รับการตรวจการทำงานของตับซึ่งแบ่งเป็น 2 กลุ่ม คือ กลุ่มผู้ป่วยที่มีแนวโน้มเป็น และ ไม่เป็น โรคตับ โดยศึกษาเปรียบเทียบเทคนิคการวิเคราะห์ทางสถิติ 2 วิธี คือ การวิเคราะห์จำแนกกลุ่ม และการวิเคราะห์ถดถอยโลจิสติก เกณฑ์ในการเปรียบเทียบคือ ค่าอัตราพยากรณ์ที่ผิดพลาด (Apparent Error Rate: APER) โดยวิธีที่มีประสิทธิภาพสูงสุดคือ วิธีที่มีค่าอัตราพยากรณ์ที่ผิดพลาดต่ำสุด ซึ่งตัวแปรอิสระที่ศึกษาในงานวิจัยนี้มี 8 ตัว คือ Total protein (TP: x_1), Albumin (Alb: x_2), Globulin (Glb: x_3), Total Bilirubin (TB: x_4), Direct Bilirubin (DB: x_5), SGOT (x_6), SGPT (x_7) และ Alkaline Phosphatase (ALP: x_8) ซึ่งผลการศึกษสามารถสรุปได้ ดังนี้

ผลการวิเคราะห์จำแนกกลุ่ม พบว่า ฟังก์ชันที่ใช้ในการจำแนกกลุ่มคือ

$$\begin{aligned} \hat{Y}_1 = & -109.090 + 12.996x_1 + 11.447x_2 + 12.969x_3 + 16.128x_4 - 4.460x_5 + 0.238x_6 - 0.015x_7 + 0.245x_8 \\ & + 105.718x_1x_2 + 105.896x_1x_3 + 0.158x_1x_4 - 1.152x_1x_5 + 0.114x_1x_6 - 0.052x_1x_7 - 0.018x_1x_8 - 103.452x_2x_3 - 0.350x_2x_4 \\ & - 3.432x_2x_5 - 0.154x_2x_6 + 0.102x_2x_7 - 0.002x_2x_8 - 1.702x_3x_4 + 6.850x_3x_5 - 0.050x_3x_6 + 0.002x_3x_7 + 0.018x_3x_8 \\ & + 13.890x_4x_5 + 0.048x_4x_6 + 0.018x_4x_7 - 0.040x_4x_8 - 0.158x_5x_6 + 0.006x_5x_7 + 0.126x_5x_8 + 0.006x_6x_7 - 53.741x_7^2 \\ & - 54.977x_8^2 - 56.769x_1^2 - 8.797x_2^2 - 51.671x_3^2 - 0.008x_4^2 - 0.005x_5^2 - 0.001x_6^2 \\ \hat{Y}_2 = & -27.429 - 1.910x_1 + 6.803x_2 + 6.263x_3 + 1.581x_4 - 0.866x_5 + 0.018x_6 - 0.006x_7 + 0.006x_8 + 112.122x_1x_2 \\ & + 112.296x_1x_3 + 0.608x_1x_4 - 0.778x_1x_5 + 0.012x_1x_6 - 0.002x_1x_7 + 0.002x_1x_8 - 112.030x_2x_3 - 0.306x_2x_4 + 0.060x_2x_5 \\ & - 0.014x_2x_6 + 0.006x_2x_7 - 0.002x_2x_8 - 0.484x_3x_4 + 0.612x_3x_5 - 0.006x_3x_6 - 0.002x_3x_7 + 0.002x_3x_8 + 3.392x_4x_5 - 0.006x_4x_6 \\ & + 0.002x_4x_7 - 0.004x_4x_8 + 0.014x_5x_6 - 0.004x_5x_7 + 0.006x_5x_8 - 56.158x_6^2 - 56.767x_7^2 - 56.947x_8^2 - 1.413x_1^2 - 2.460x_2^2 \end{aligned}$$

โดยมีความถูกต้องแม่นยำในการพยากรณ์ 97.67% หรือมีความคลาดเคลื่อนในการพยากรณ์ประมาณ 2.3%

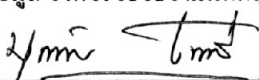
ผลการวิเคราะห์การถดถอยแบบโลจิสติก พบว่า สมการถดถอยคือ

$$\hat{g}(x) = 13.2738 + 0.1503(x_1 + x_2 + x_3) - 3.7401x_4 + 0.9760x_5 - 0.1154x_6 + 0.00787x_7 - 0.0466x_8$$

ซึ่งเมื่อทดสอบความเหมาะสมของตัวแบบแล้วพบว่า ไม่เหมาะสมกับข้อมูล แม้ว่าจะทำให้ความถูกต้องแม่นยำในการพยากรณ์ 94.33% หรือมีความคลาดเคลื่อนในการพยากรณ์ประมาณ 5.7% ก็ตาม

ผลการเปรียบเทียบอัตราความผิดพลาดในการพยากรณ์การจำแนกกลุ่มผู้ป่วยโรคตับ พบว่า การวิเคราะห์จำแนกกลุ่มมีประสิทธิภาพสูงกว่าวิธีการถดถอยโลจิสติก และเนื่องจากการวิเคราะห์การถดถอยโลจิสติกนั้น ตัวแบบที่ได้ไม่เหมาะสมกับลักษณะข้อมูล จึงควรใช้วิธีจำแนกกลุ่มสำหรับข้อมูลชุดนี้

ศยามล ลำลองรัตน์
ลายมือชื่อนิติ


ลายมือชื่อประธานกรรมการ

15 / 10 / 2550