Syamol Lumlongrut 2007: A Comparative Study of Liver Disease Diagnosis Using Discriminant Analysis and Logistic Regression Analysis for the Patients in Chachoengsao Hospital. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Boonorm Chomtee, Ph.D. 91 pages.

The purpose of this research was to compare the two statistical methods : discriminant analysis and logistic regression analysis in liver diagnosis for the patients in Chachoengsao hospital. These patients who were taken the liver function test will be classified to 2 groups : a group of patients who had possibility to get liver disease and a group of patients who had not possibility to got liver disease using the two statistical methods and the criteria for comparison is the apparent error rate(APER) which the lower APER is the higher efficiency method. The liver function test data are 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) and Alkaline Phosphatase (ALP: x_8). The results of the two statistical methods are as follows :

For discriminant analysis, the quadratic function is

$\hat{Y}_1 = -109.090+12.996x_1+11.447x_2+12.969x_3+16.128x_4-4.460x_5+0.238x_6-0.015x_7+0.245x_8$
$+105.718x_1x_2+105.896x_1x_3+0.158x_1x_4-1.152x_1x_5+0.114x_1x_6-0.052x_1x_7-0.018x_1x_8-103.452x_2x_3-0.350x_2x_4$
$-3.432x_2x_5-0.154x_2x_6+0.102x_2x_7-0.002x_2x_8-1.702x_3x_4+6.850x_3x_5-0.050x_3x_6+0.002x_3x_7+0.018x_3x_8$
$+13.890x_4x_5+0.048x_4x_6+0.018x_4x_7-0.040x_4x_8-0.158 x_5x_6+0.006x_5x_7+0.126x_5x_8+0.006x_6x_7-53.741x_1^2$
$-54.977x_2^2-56.769x_3^2-8.797x_4^2-51.671x_5^2-0.008x_6^2-0.005x_7^2-0.001x_8^2.$

$\hat{Y}_2 = -27.429-1.910x_1+6.803x_2+6.263x_3+1.581x_4-0.866x_5+0.018x_6-0.006x_7+0.006x_8+112.122x_1x_2$
$+112.296x_1x_3+0.608x_1x_4-0.778x_1x_5+0.012x_1x_6-0.002x_1x_7+0.002x_1x_8-112.030x_2x_3-0.306x_2x_4+0.060x_2x_5$
$-0.014x_2x_6+0.006x_2x_7-0.002x_2x_8-0.484x_3x_4+0.612x_3x_5-0.006x_3x_6-0.002x_3x_7+0.002x_3x_8+3.392x_4x_5-0.006x_4x_6$
$+0.002x_4x_7-0.004x_4x_8+0.014x_5x_6-0.004x_5x_7+0.006x_5x_8-56.158x_1^2-56.767x_2^2-56.947x_3^2-1.413x_4^2-2.460x_5^2.$

The approximate accuracy of the function is 97.67%.

For logistic regression analysis, the regression model is

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

However , after testing the logistic regression analysis, it was found that regression model is not appropriate at $\alpha = 0.05$.

For the comparison between these two statistical methods, it is shown that discriminant analysis is more efficiency than logistic regression analysis based on APER. Moreover, discriminant analysis was recommended because of the inappropriate of the logistic regression model.

_____   _____   ___ / ___ / ___

Student's signature   Thesis Advisor's signature