Logistic Model Performance Evaluation for Liver Patients Data

Background Information:

(1) Patients Suffering from Liver disease have been continuously increasing because of excessive consumption of alcohol, inhale of harmful gases, intake of contaminated food and drugs.

Problems with liver patients are not easily discovered in an early stage, as the liver will be functioning normally even when it is partially damaged.

Liver disease can be diagnosed by analyzing the levels of enzymes in the blood. An early diagnosis of liver problems will help to increase the patient' s survival rate.

(2) Data Set details:

Age	Gender	TB	DB	Alkphos	Sgpt	Sgot	TP	ALB	AGratio	Target
7	Female	27.2	11.8	1420	790	1050	6.1	2	0.4	1
19	Male	1.4	0.8	178	13	26	8	4.6	1.3	2

(Data Set Sample)

The data set contains 416 liver patient records and 167 non-liver patients. Total is 583 record.

The data set has a total of 11 attributes. Out of which 10 are Independent attributes and 1 is the Target attribute.

(3) Attributes details

1. Age: Age of the patient
2. Gender: Gender of the patient
3. TB: Total Bilirubin
4. DB: Direct Bilirubin
5. Alkphos: Alkaline Phosphotase
6. Sgpt: Alamine Aminotransferase
7. Sgot: Aspartate Aminotransferase
8. TP: Total Protiens
9. ALB: Albumin
10. A/G: Ratio Albumin and Globulin Ratio
11. Selector field used to split the data into two sets (labeled by the experts) (1 - indicates liver patients, 2 - indicates non liver patients)

(4) Data Division:

Training data set is used to develop the model. 478 records are used.

Testing data used to evaluate the model developed using the Training data set. 105 records are used.

Investigation:

1. Data clean:

```
/*Clean Data*/
data patients;
    set liver_patients;
    if Gender='Female' then GenderGroup=1;
    else GenderGroup=0;

run;
/*Clean Data*/
data test_patients;
    set test;
    if Gender='Female' then GenderGroup=1;
    else GenderGroup=0;

run;
```

Change 'Female' to 1 and 'Male' to 0. And build a new column, named GenderGroup.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.1391	0.8457	6.3976	0.0114
Age	1	0.0202	0.00682	8.8005	0.0030
GenderGroup	1	-0.0274	0.2452	0.0124	0.9112
TB	1	0.0138	0.0799	0.0298	0.8629
DB	1	0.2902	0.2236	1.6842	0.1944
Alkphos	1	0.00127	0.000834	2.3257	0.1273
Sgpt	1	0.00949	0.00511	3.4558	0.0630
Sgot	1	0.00271	0.00320	0.7142	0.3980
TP	1	0.5096	0.1983	6.6059	0.0102
ALB	1	-0.7185	0.2813	6.5241	0.0106

2. Logistic Model:

logit(p(liver_patient)) = log(p/1-p)

=-2.1391 + 0.0202*Age - 0.0274*Gender + 0.0138*TB + 0.2902*DB + 0.00127*Alkphos + 0.00949*Sgpt + 0.00271*Sgot + 0.5096*TP - 0.7185*ALB; #Gender == GenderGroup

The formula for liver patient probability: p = exp(logit)/(exp(logit) + 1)

```
data test_p;
    set test_patients;
    logit = -2.1391 + 0.0202*Age - 0.0274*GenderGroup + 0.0138*TB + 0.2902*DB +
    0.00127*Alkphos + 0.00949*Sgpt + 0.00271*Sgot + 0.5096*TP - 0.7185*ALB;

    p=exp(logit)/(exp(logit)+1);
    if p < 0.5 then Target_predicted = '2';
    else Target_predicted = '1';
    keep Target p Target_predicted;

run;
```

Substitute logic formula into 'Testing Data Set', if threshold p less than 0.5, the guy predicted to be the liver patient, else he or she is non liver patient. But from the table of part of sample result 'test_p',

Obs	Target	p	Target_predicted
1	1	0.99970	1
2	1	0.99997	1
3	2	0.47436	2
4	1	0.93296	1
5	1	0.79082	1
6	2	0.56057	1
7	1	0.77572	1
8	1	0.97491	1
9	1	0.61513	1

(test_p result table)

I find, people who is non liver patient predicted to be the liver patient.

3. Evaluation

In order to evaluate this logistic regression:

The first step is build a 'Confusion Matrix'. A **confusion matrix** is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

Confusion Table

The FREQ Procedure

Frequency	Table of Target by Target_predicted		
	Target_predicted		
Target	1	2	Total
1	78	3	81
2	20	4	24
Total	98	7	105

(confusion matrix)

I. Accuracy VS Error Rate:

Accuracy = true positive and true negative / total cases = (78 + 4) / 105 = 78.1%

Error Rate = false positive and false negative / total cases = (20 + 3) / 105 = 21.9%

II. Recall VS PV+:

Recall (True Positive Rate/ Sensitivity) = true positive/ total actual positive = 78 / 81 = 96.30% (TPR)

Precision(Positive Predicted Value , PV+) = true positive / total predicted positive = 78 / 98 = 79.59%

III Specificity VS PV-:

Specificity(True Negative Rate) = true negative / total actual negative = 4 / 24 = 16.67%

Negative Predicted Value(PV-) = true negative / total predicted negative = 4 / 7 = 57.14%

All of these data can deliver from **proc freq procedure**.

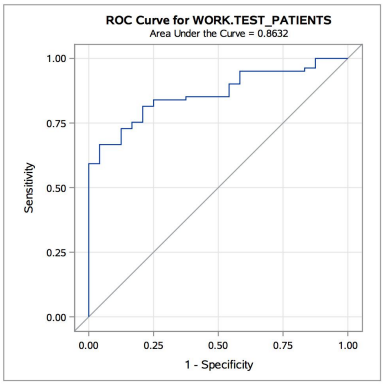
The FREQ Procedure

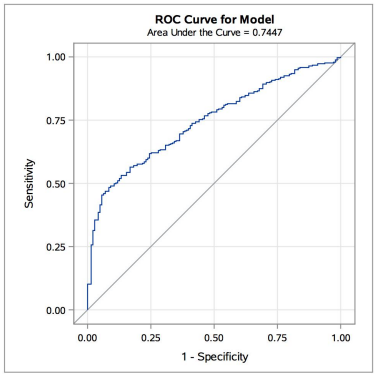
Frequency Percent Row Pct Col Pct	Table of Target by Target_predicted		
	Target_predicted		
Target	1	2	Total
1	78 74.29 96.30 79.59	3 2.86 3.70 42.86	81 77.14
2	20 19.05 83.33 20.41	4 3.81 16.67 57.14	24 22.86
Total	98 93.33	7 6.67	105 100.00

IV F1-measure:

F1-measure = 2*true positive / 2*true positive + negative positive +actual negative = (2 * 78) / (78 + 98 + 20) = 79.59%

Second Step: ROC





Compare to the AUC equals to 0.7447 in train data, the test data's AUC is 0.8632, bigger than train data's. It shows this logistic model would be considered to be 'good' at classifying people is liver patient or not.