Enhancing the Output from PROC LOGISTIC

Bharat Thakkar, Kwan Hur, Charles A. Oprian, William G. Henderson, and Sharon Urbanski
Center for Cooperative Studies in Health Services, Department of Veterans Affairs Hospital, Hines, IL

ABSTRACT

A SAS®code for enhancing the output from PROC LOGISTIC with the stepwise option was developed. Among the enhancements provided are: the suppression of SAS output from several pages to one table, the incremental change in the c-index statistic, the -2 LOG likelihood statistic (-2 log L statistic), and an incremental test for the significant change by using the -2 log L statistic. In addition, the code provides the parameter estimate, the standard error, the Wald chi-square, the p-value of the chi-square, the standardized estimate, and the odds ratio at each step of the stepwise logistic regression output.

INTRODUCTION

One aspect of data analysis is to find the relationship between a response variable and a set of explanatory variables. When the response variable is dichotomous, logistic regression is used to build a model. The stepwise selection procedure provides an effective means to screen a large number of explanatory variables, and to simultaneously fit a number of logistic regression models.

The c-index is one of the statistics that can be used to measure the predictability of the logistic regression model. Another statistic that is used to evaluate a regression model is the -2 log L statistic. This information is presented in a single table of the enhanced output.

EXAMPLE

In the following example a subset of 2,056 surgical cases that had a colectomy for colon cancer was identified, from the National VA Surgical Quality Improvement Program which contained 107,241 major surgery cases performed in the VA system, with 30 day mortality as the response variable and 15 selected risk factors as explanatory variables.

```
PROC LOGISTIC DATA=SAMPLE ORDER=FORMATTED;
  MODEL death= dyspnea dnr fnstatus hxcopd cpneumon
               hepatomg ascites impsens hxtia cvaneuro
               discancr wtloss bleeddis transfus emergncy /
      selection = stepwise slstay=0.05 slentry=0.05 details;
```

Note: *slentry*-the significance level for entry into the model, selected to be 0.05, *slstay*-the significance level for staying in the model, selected to be 0.05, and the *details* produce "analysis of variables NOT in the model" and "analysis of maximum likelihood estimates" tables at each step. After submitting the above SAS code, the following seven page standard SAS PROC LOGISTIC® output is produced.

********** SAS output **********

```
The SAS System                                          Pg. 1

          The LOGISTIC Procedure

Data Set: IN.LONGO
Response Variable: DEATH
Response Levels: 2
Number of Observations: 2056
Link Function: Logit

             Response Profile
Ordered
Value  DEATH        Count

    1  Dead           106
    2  Not Dead      1950

Stepwise Selection Procedure

Step  0. Intercept entered:

     Analysis of Maximum Likelihood Estimates

                Parameter  Standard  Wald        Pr >        Standardized  Odds
Variable  DF      Estimate     Error  Chi-Square  Chi-Square  Estimate      Ratio
INTERCPT  1      -2.9121    0.0997  852.5965    0.0001      .             0.054

     Residual Chi-Square = 145.8735 with 15 DF (p=0.0001)

          Analysis of Variables Not in the Model

                          Score          Pr >
Variable                  Chi-Square     Chi-Square
DYSPNEA                   4.8542         0.0276
DNR                       4.8391         0.0278
FNSTATUS                  9.8128         0.0017
HXCOPD                    2.5937         0.1073
CPNEUMON                  6.0945         0.0136
HEPATOMG                  18.5748        0.0001
ASCITES                   50.1471        0.0001
IMPSENS                   41.9061        0.0001
```

HXTIA	2.5009	0.1138
CVANEURO	1.5489	0.2133
DISCANCR	30.0863	0.0001
WTLOSS	9.6719	0.0019
BLEEDDIS	6.6133	0.0101
TRANSFUS	13.8493	0.0002
EMERGNCY	9.9439	0.0016

The LOGISTIC Procedure

Step 1. Variable ASCITES entered:

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	837.035	816.746	.
SC	842.664	828.003	.
-2 LOG L	835.035	812.746	22.290 with 1 DF (p=0.0001)
Score	.	.	50.147 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-2.9844	0.1035	830.8623	0.0001	.	0.051
ASCITES	1	2.5790	0.4680	30.3631	0.0001	0.139587	13.184

Association of Predicted Probabilities and Observed Responses

Concordant	= 7.5%	Somers' D	= 0.069
Discordant	= 0.6%	Gamma	= 0.859
Tied	= 91.9%	Tau-a	= 0.007
(206700 pairs)		c	= 0.535

Residual Chi-Square = 92.5433 with 14 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
DYSPNEA	3.9366	0.0472
DNR	5.5599	0.0184
FNSTATUS	10.6446	0.0011
HXCOPD	2.7344	0.0982
CPNEUMON	2.1787	0.1399
HEPATOMG	4.2202	0.0399
IMPSENS	39.2517	0.0001
HXTIA	2.8812	0.0896
CVANEURO	2.1475	0.1428
DISCANCR	24.9037	0.0001
WTLOSS	8.6668	0.0032
BLEEDDIS	5.7791	0.0162
TRANSFUS	10.3510	0.0013
EMERGNCY	9.2530	0.0024

The LOGISTIC Procedure

Step 2. Variable IMPSENS entered:

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	837.035	793.227	.
SC	842.664	810.113	.
-2 LOG L	835.035	787.227	47.808 with 2 DF (p=0.0001)
Score	.	.	89.904 with 2 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.1395	0.1136	763.7191	0.0001	.	0.043
ASCITES	1	2.5724	0.4813	28.5649	0.0001	0.139233	13.098
IMPSENS	1	1.6269	0.2834	32.9560	0.0001	0.192076	5.088

Association of Predicted Probabilities and Observed Responses

Concordant	= 23.7%	Somers' D	= 0.201
Discordant	= 3.6%	Gamma	= 0.734
Tied	= 72.7%	Tau-a	= 0.020
(206700 pairs)		c	= 0.600

Residual Chi-Square = 49.5609 with 13 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
DYSPNEA	2.7828	0.0953
DNR	1.0269	0.3109
FNSTATUS	0.3701	0.5430
HXCOPD	2.0544	0.1518
CPNEUMON	1.7011	0.1921
HEPATOMG	3.7391	0.0532
HXTIA	3.8051	0.0511
CVANEURO	0.0826	0.7738
DISCANCR	22.4806	0.0001
WTLOSS	7.3106	0.0069
BLEEDDIS	4.1473	0.0417
TRANSFUS	7.5478	0.0060
EMERGNCY	7.2757	0.0070

The LOGISTIC Procedure

Step 3. Variable DISCANCR entered:

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	837.035	777.296	.
SC	842.664	799.810	.
-2 LOG L	835.035	769.296	65.739 with 3 DF (p=0.0001)
Score	.	.	113.385 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio

INTERCPT	1	-3.3160	0.1270	681.3329	0.0001	.	0.036
ASCITES	1	2.4052	0.4977	23.3530	0.0001	0.130181	11.081
IMPSENS	1	1.5917	0.2882	30.5099	0.0001	0.187917	4.912
DISCANCR	1	1.1116	0.2433	20.8748	0.0001	0.187962	3.039

Association of Predicted Probabilities and Observed Responses

Concordant	= 39.3%	Somers' D	= 0.304
Discordant	= 8.9%	Gamma	= 0.631
Tied	= 51.7%	Tau-a	= 0.030
(206700 pairs)		c	= 0.652

Residual Chi-Square = 26.7569 with 12 DF (p=0.0084)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
DYSPNEA	2.6159	0.1058
DNR	0.7627	0.3825
FNSTATUS	0.6833	0.4085
HXCOPD	2.2454	0.1340
CPNEUMON	1.4709	0.2252
HEPATOMG	3.1590	0.0755
HXTIA	3.3915	0.0655
CVANEURO	0.3677	0.5442
WTLOSS	4.4180	0.0356
BLEEDDIS	2.3061	0.1289
TRANSFUS	5.4018	0.0201
EMERGNCY	8.5698	0.0034

The LOGISTIC Procedure

Step 4. Variable EMERGNCY entered:

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	837.035	772.386	.
SC	842.664	800.529	.
-2 LOG L	835.035	762.386	72.649 with 4 DF (p=0.0001)
Score	.	.	121.563 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.3998	0.1339	644.7836	0.0001	.	0.033
ASCITES	1	2.3825	0.5032	22.4129	0.0001	0.128952	10.832
IMPSENS	1	1.5518	0.2889	28.8511	0.0001	0.183206	4.720
DISCANCR	1	1.1472	0.2446	22.0065	0.0001	0.193992	3.149
EMERGNCY	1	0.9049	0.3174	8.1299	0.0044	0.119247	2.472

Association of Predicted Probabilities and Observed Responses

Concordant	= 45.7%	Somers' D	= 0.342
Discordant	= 11.5%	Gamma	= 0.599
Tied	= 42.9%	Tau-a	= 0.033
(206700 pairs)		c	= 0.671

Residual Chi-Square = 18.0431 with 11 DF (p=0.0806)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
DYSPNEA	2.1573	0.1419
DNR	0.4120	0.5209
FNSTATUS	0.3043	0.5812
HXCOPD	2.0851	0.1487
CPNEUMON	1.2288	0.2676
HEPATOMG	3.9719	0.0463
HXTIA	3.1348	0.0766
CVANEURO	0.4170	0.5184
WTLOSS	4.1854	0.0408
BLEEDDIS	1.8481	0.1740
TRANSFUS	4.6738	0.0306

The LOGISTIC Procedure

Step 5. Variable TRANSFUS entered:

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	837.035	770.537	.
SC	842.664	804.308	.
-2 LOG L	835.035	758.537	76.498 with 5 DF (p=0.0001)
Score	.	.	128.554 with 5 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.4217	0.1348	644.2881	0.0001	.	0.033
ASCITES	1	2.3163	0.5116	20.4974	0.0001	0.125369	10.138
IMPSENS	1	1.5161	0.2902	27.2857	0.0001	0.178992	4.554
DISCANCR	1	1.1050	0.2465	20.0984	0.0001	0.186852	3.019
TRANSFUS	1	0.9354	0.4418	4.4824	0.0342	0.077095	2.548
EMERGNCY	1	0.8783	0.3207	7.4982	0.0062	0.115737	2.407

Association of Predicted Probabilities and Observed Responses

Concordant	= 48.0%	Somers' D	= 0.359
Discordant	= 12.1%	Gamma	= 0.597
Tied	= 39.8%	Tau-a	= 0.035
(206700 pairs)		c	= 0.680

Residual Chi-Square = 13.3137 with 10 DF (p=0.2067)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
DYSPNEA	1.5497	0.2132
DNR	0.4538	0.5006
FNSTATUS	0.0791	0.7786
HXCOPD	2.0449	0.1527
CPNEUMON	0.5155	0.4728
HEPATOMG	3.1570	0.0756
HXTIA	3.0383	0.0813
CVANEURO	0.2909	0.5897
WTLOSS	3.3336	0.0679
BLEEDDIS	0.8911	0.3452

The LOGISTIC Procedure

NOTE: No (additional) variables met the 0.05 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Removed	Number In	Score Chi-Square	Wald Chi-Square	Pr > Chi-Square
1	ASCITES		1	50.1471	.	0.0001
2	IMPSENS		2	39.2517	.	0.0001
3	DISCANCR		3	22.4806	.	0.0001
4	EMERGNCY		4	8.5698	.	0.0034
5	TRANSFUS		5	4.6738	.	0.0306

********** end of SAS output **********;

In order to reformat the above SAS output file, the SAS code used requires SAS/BASE®, SAS/STAT® software (SAS/GRAPH® software is optional) and works on all platforms and can be obtained by contacting the author. The first step is to read the SAS file using a SAS DATA step along with the SAS statement INFILE. Using the DO While loop and the INPUT statement (INPUT @1 rec $char131.;) to read each record of the file.

The first task is to stop the cursor at each record containing ('Step ') using the INDEX function. In other words, the cursor will stop at the record 'Step 0. Intercept entered:' first. With the cursor stopped, read the step number, the variable name, and entered or removed status using the INDEX function and the SUBSTR function. Determine if the current record is the initial Step and output to a SAS dataset.

The second task depends on if the cursor is at the initial step. If this is true then read the "Analysis of maximum likelihood estimate" table for parameter estimate, standard error, wald chi-square, pr> chi-square, standardized estimate and the odds ratio for each of the variables listed. First, move the cursor using the INDEX function to the record containing the text 'INTERCPT', which is always the first record of the table. To determine the number of records in the table, add one to the step number(ex. for the initial step number of records= 0 + 1). Now set up a DO loop and read the table using the SUBSTR function along with the INPUT statement. After determining that the cursor is NOT at the initial step, than determine if the step is entering or removing the variable. If the variable is removed then output to a SAS dataset and move cursor to the next step and continue. After determining that the variable is entered into the model, read the criteria for assessing model fit table for the - 2 log L statistic and the score statistic along with the appropriate p-values using similar steps as above.

Next, move the cursor to the analysis of maximum likelihood estimates table and read the similar information as above and output to a SAS dataset.

Finally, move the cursor to the association of predicted probabilities and observed responses table and read the c-index statistic and output to a SAS dataset. Repeat the above steps for all the steps in the output file.

Using the DATA _NULL_ step with the appropriate put statements a SAS code creates table 1 and creates the logit equation. Based on the information from table1 and SAS/GRAPH® software(or use any other graphical software or use PROC PLOT®) creates figure 1 if desired.

The above step can be applied to any SAS procedural output files. Some useful applications include reading the SAS PROC FREQ® output file or the SAS PROC TTEST® output file for univariate screening for statistical model building.

The enhanced Logistic output compresses the seven page output into a single table (Table 1) which includes the c-index, -2 log L statistic, score statistic with the corresponding p-values from each of the five steps, and also provides statistical information (standard error, wald chi-square, odds ratio, standardized estimate, pr > chi-square, parameter estimate, and explanatory variables) for each of the steps. In addition, the output shows a test for significant change for each of the steps. The test compares the -2 log L statistic for the previous and current step (delta -2 log L statistic). Statistical significant can be checked using chi-square 1 degree of freedom. The logit equation of the probability is

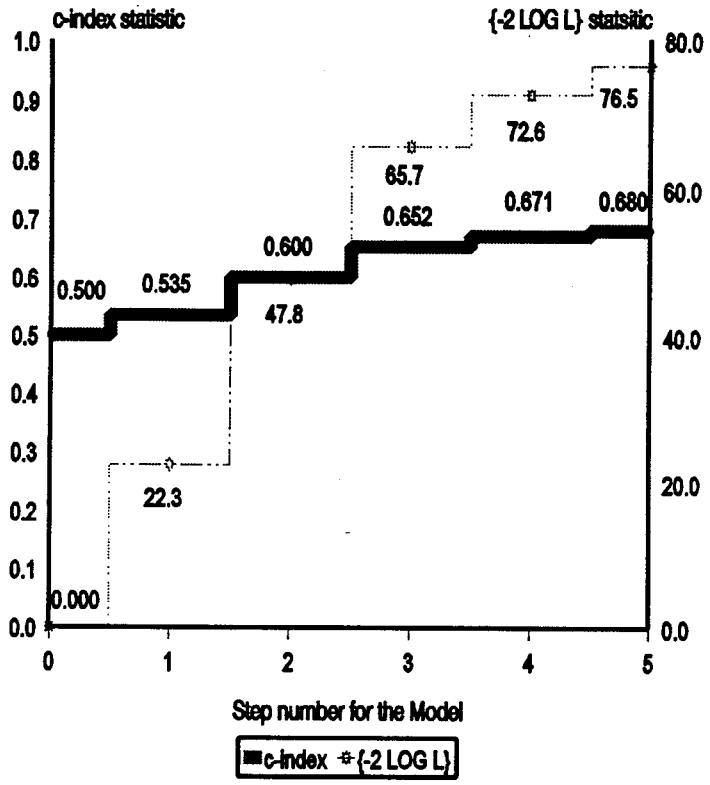

Figure 1. Change in c-index & -2 log L statistic

available at each step. For example, the logit equation for step # 4 is

*Logit(p) = -3.3998 + (2.3825) * Ascites + (1.5518) * Imspens + (1.1472)*Discancr +(0.9049) * Emergncy*

CONCLUSION

The enhanced SAS output provides information about the relationship between the response & explanatory variables in a more comprehensible format than the standard SAS PROC LOGISTIC® output. In addition, this enhanced output can provide the incremental change in the c-index and the -2 log L statistic. This enhanced output becomes more useful as the number of explanatory variables (and steps) increases. Finally figure 1 shows the change in the c-index (delta c-index is depicted as the vertical solid lines) and the change in the -2 log L statistic (delta -2 log L statistic is depicted as the vertical dashed lines).

REFERENCES

1. SAS Institute Inc. (1990), "The LOGISTIC Procedure," Chapter 27 of SAS/STAT User's Guide Version 6, Fourth Edition, Volume 2, Cary, NC SAS Institute Inc.

2. Hosmer, DW and Lemeshow, S. Applied Logistic Regression. New York: John Wiley Sons, 1989.

3. Iezzoni, Lisa (EV) Risk Adjustment for ModernHealth Care Outcomes Health Administration Press Ann Arbor, Mi 1994.

4. Owen, D. B., Handbook of Statistical Tables Addison-Wesley Reading, Mass, 1962 pp50-55.

ACKNOWLEDGMENTS

I wish to thank the National VA Surgical Quality Improvement Program, for allowing us to use a portion of the NSQIP data. This material is based upon work supported by the Health Services Research and Development Services, Department of Veterans Affairs.

AUTHOR CONTACT INFORMATION

Department of Veterans Affairs, Hines CCSHS
Bharat Thakkar (151A) Rm. 1B249
Fifth Avenue & Roosevelt Road Bldg. 1
Hines, Il 60141 (708) 343-7200 x 3555

Steps	Explanatory Variables	Parameter Estimate	Standard Error	Wald Chi-sq.	Pr > Chi-sq.	Standardized Estimate	Odds Ratio	c_Index	Delta c_Index	Delta (-2LOG L)	Delta (-2 LOG L)	Significant Change	p-value	Score	p-value
0	INTERCPT	-2.9121	0.0997	852.5965	0.0001		0.054	0.500			0.000				
1	INTERCPT	-2.9844	0.1035	830.8623	0.0001		0.051	0.535	0.035	22.290	22.290	Y	0.0001	50.147	0.0001
	ASCITES	2.5790	0.4680	30.3631	0.0001	0.1396	13.184								
2	INTERCPT	-3.1395	0.1136	763.7191	0.0001		0.043	0.600	0.065	47.808	25.518	Y	0.0001	89.904	0.0001
	ASCITES	2.5724	0.4813	28.5649	0.0001	0.1392	13.098								
	IMPSENS	1.6269	0.2834	32.9560	0.0001	0.1921	5.088								
3	INTERCPT	-3.3160	0.1270	681.3329	0.0001		0.036	0.652	0.052	65.739	17.931	Y	0.0001	113.385	0.0001
	ASCITES	2.4052	0.4977	23.3530	0.0001	0.1302	11.081								
	IMPSENS	1.5917	0.2882	30.5099	0.0001	0.1879	4.912								
	DISCANCR	1.1116	0.2433	20.8748	0.0001	0.1880	3.039								
4	INTERCPT	-3.3998	0.1339	644.7836	0.0001		0.033	0.671	0.019	72.649	6.910	Y	0.0001	121.563	0.0001
	ASCITES	2.3825	0.5032	22.4129	0.0001	0.1290	10.832								
	IMPSENS	1.5518	0.2889	28.8511	0.0001	0.1832	4.720								
	DISCANCR	1.1472	0.2446	22.0065	0.0001	0.1940	3.149								
	EMERGNCY	0.9049	0.3174	8.1299	0.0044	0.1192	2.472								
5	INTERCPT	-3.4217	0.1348	644.2881	0.0001		0.033	0.680	0.009	76.498	3.849	Y	0.0001	128.554	0.0001
	ASCITES	2.3163	0.5116	20.4974	0.0001	0.1254	10.138								
	IMPSENS	1.5161	0.2902	27.2857	0.0001	0.1790	4.554								
	DISCANCR	1.1050	0.2465	20.0984	0.0001	0.1869	3.019								
	TRANSFUS	0.9354	0.4418	4.4824	0.0342	0.0771	2.548								
	EMERGNCY	0.8783	0.3207	7.4982	0.0062	0.1157	2.407								

Table 1. Enhanced Logistic Output