The CONTENTS Procedure

Data Set Name	WORK.TEMP1	**Observations**	2006
Member Type	DATA	**Variables**	29
Engine	V9	**Indexes**	0
Created	04/19/2024 10:50:54	**Observation Length**	272
Last Modified	04/19/2024 10:50:54	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	5
First Data Page	1
Max Obs per Page	481
Obs in First Data Page	460
Number of Data Set Repairs	0
Filename	/saswork/SAS_work6E9C00005CB6_odaws02-usw2.oda.sas.com/SAS_workFD0100005CB6_odaws02-usw2.oda.sas.com/temp1.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2847051
Access Permission	rw-r--r--
Owner Name	jacktubbs
File Size	768KB
File Size (bytes)	786432

The CONTENTS Procedure

Variables in Creation Order						
#	**Variable**	**Type**	**Len**	**Format**	**Informat**	**Label**
1	ID1	Num	8	BEST.		ID1
2	UNIT	Char	30	$30.	$30.	UNIT
3	Job_Description	Char	25	$25.	$25.	Job Description
4	AGE	Num	8	BEST.		AGE
5	Gender	Char	6	$6.	$6.	Gender
6	Height	Num	8	COMMA15.1		Height
7	BP	Char	7	$7.	$7.	BP
8	HTN	Num	8	BEST.		HTN
9	HR	Num	8	BEST.		HR
10	BMI	Num	8	COMMA15.1		BMI
11	Calcium_Score	Num	8	COMMA15.2		Calcium Score
12	TC	Num	8	BEST.		TC
13	LDL	Num	8	BEST.		LDL
14	HDL	Num	8	BEST.		HDL
15	TG	Num	8	BEST.		TG
16	ApoB	Num	8	BEST.		ApoB
17	sdLDL_C	Num	8	BEST.		sdLDL-C
18	Lp_a_	Num	8	BEST.		Lp(a)
19	hc_CRP	Num	8	COMMA15.1		hc-CRP
20	LpPLA2	Num	8	BEST.		LpPLA2
21	PLA2_Out_of_Range	Num	8	BEST.		PLA2 Out of Range
22	HbA1c	Num	8	COMMA15.1		HbA1c
23	IR	Num	8	COMMA15.1		IR
24	Insulin	Num	8	BEST.		Insulin
25	Vit_D	Num	8	BEST.		Vit D
26	leo	Num	8			
27	pla2_out	Num	8			
28	n	Num	8			
29	wt	Num	8			

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
AGE	AGE	2006	45.8644317	8.8552134	22.1400000	83.0500000
ApoB	ApoB	1449	103.7536232	24.7415514	29.0000000	260.0000000
BMI	BMI	1749	37.2543739	316.8503068	0.3000000	13278.90
HDL	HDL	1970	51.1538071	14.8884867	16.0000000	187.0000000
HR	HR	1721	75.0017432	16.6890682	33.0000000	396.0000000
HbA1c	HbA1c	1929	6.2721306	17.5240122	4.4000000	537.0000000
ID1	ID1	2006	1046.50	607.3948612	1.0000000	2088.00
IR	IR	974	29.1485832	25.8753177	0	100.0000000
Insulin	Insulin	2001	11.6621489	14.4054693	0	292.0000000
LDL	LDL	1975	122.7397316	34.5311664	19.0000000	333.0000000
LpPLA2	LpPLA2	2006	149.1131605	93.8768953	0	600.0000000
TC	TC	1971	196.8538813	59.5073169	19.0000000	2216.00
TG	TG	1970	134.8304569	94.5988869	6.0000000	1520.00
hc_CRP	hc-CRP	1312	3.6293598	35.2974900	0	1270.00
sdLDL_C	sdLDL-C	1559	401.3455420	516.6419472	0.2000000	2400.00
wt		1750	205.3316000	41.8104616	103.0000000	370.0000000

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

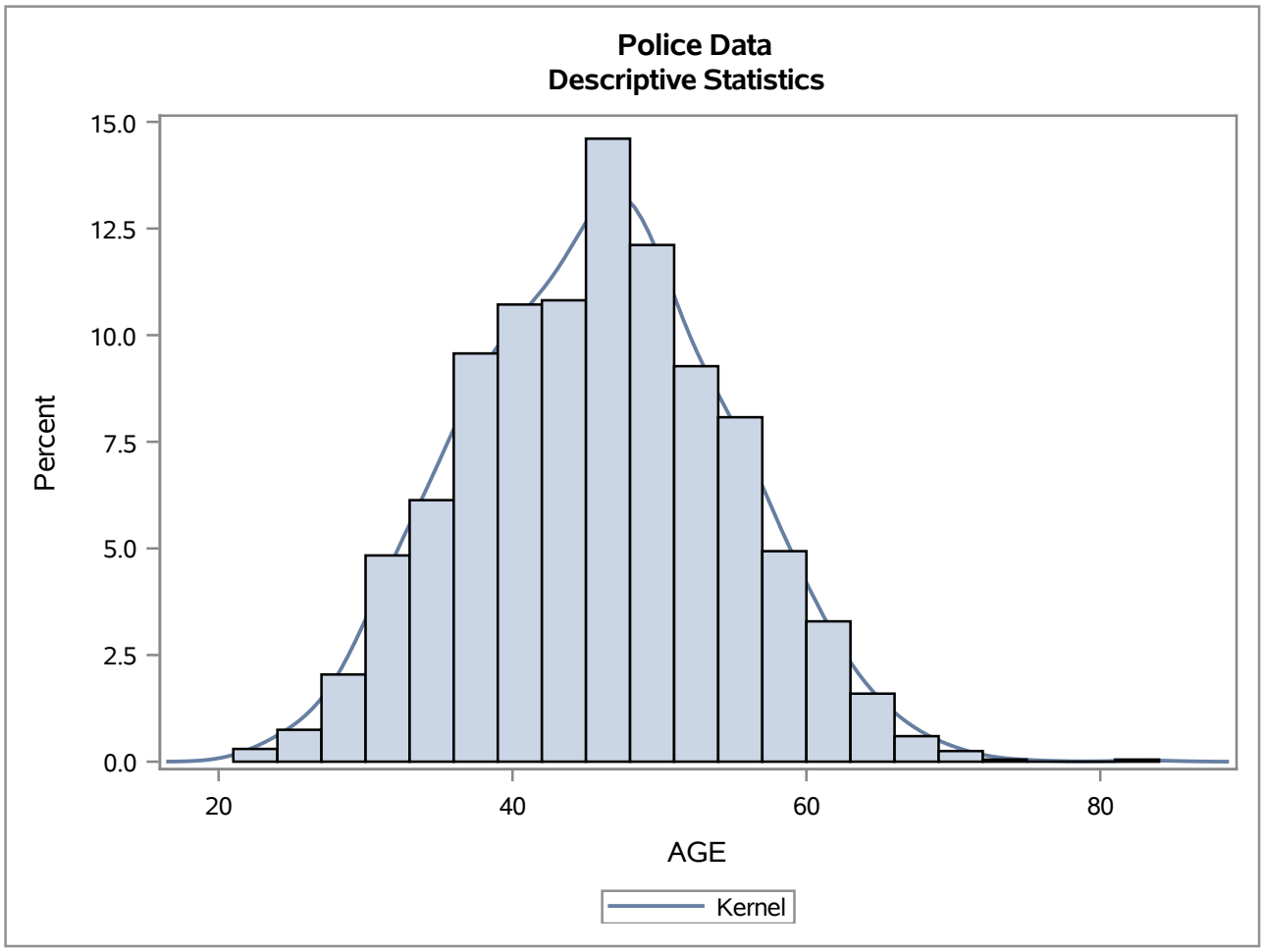
Table of leo by Gender			
	Gender(Gender)		
leo	Female	Male	Total
0	119 5.93 34.80 29.97	223 11.12 65.20 13.86	342 17.05
1	278 13.86 16.71 70.03	1386 69.09 83.29 86.14	1664 82.95
Total	397 19.79	1609 80.21	2006 100.00

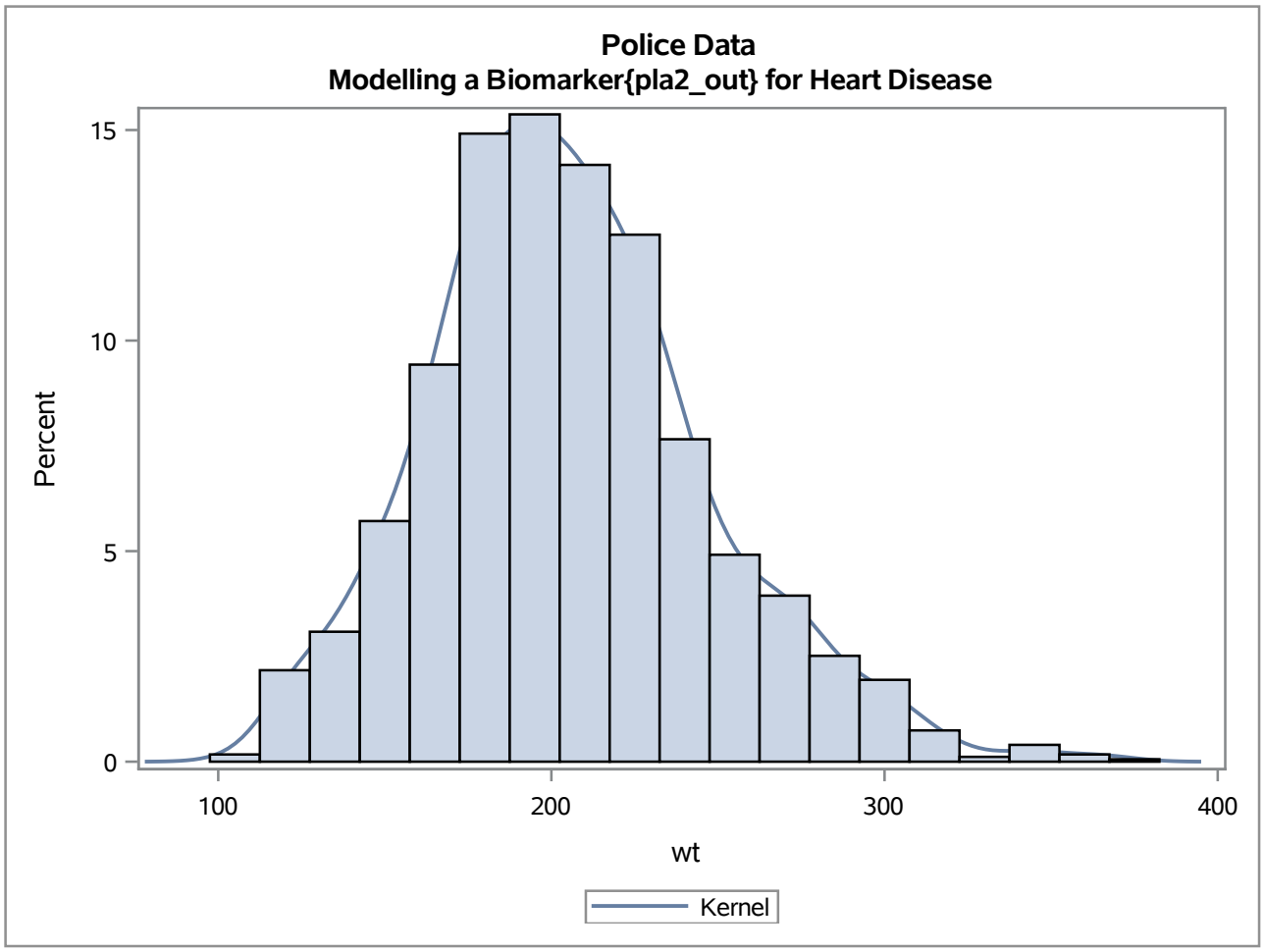
Frequency
Percent
Row Pct
Col Pct

Table of leo by pla2_out			
	pla2_out		
leo	0	1	Total
0	266 13.26 77.78 18.81	76 3.79 22.22 12.84	342 17.05
1	1148 57.23 68.99 81.19	516 25.72 31.01 87.16	1664 82.95
Total	1414 70.49	592 29.51	2006 100.00

Frequency
Percent
Row Pct
Col Pct

Table of leo by HTN			
	HTN(HTN)		
leo	0	1	Total
0	279 13.91 81.58 17.76	63 3.14 18.42 14.48	342 17.05
1	1292 64.41 77.64 82.24	372 18.54 22.36 85.52	1664 82.95
Total	1571 78.32	435 21.68	2006 100.00


Police Data
Descriptive Statistics
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Percent
20
40
60
80
AGE
Kernel


Police Data
Modelling a Biomarker{pla2_out} for Heart Disease
Percent
15
10
5
0
100
200
300
400
wt
Kernel

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEMP1
Response Variable	pla2_out
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	2006
Number of Observations Used	1975

Response Profile		
Ordered Value	**pla2_out**	**Total Frequency**
1	0	1394
2	1	581

Probability modeled is pla2_out=1.

Note: 31 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information		
Class	**Value**	**Design Variables**
Gender	**Female**	1
	Male	-1
leo	**0**	1
	1	-1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	2395.106	2210.395
SC	2400.694	2232.749
-2 Log L	2393.106	2202.395

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	190.7109	3	<.0001
Score	178.7987	3	<.0001
Wald	158.8841	3	<.0001

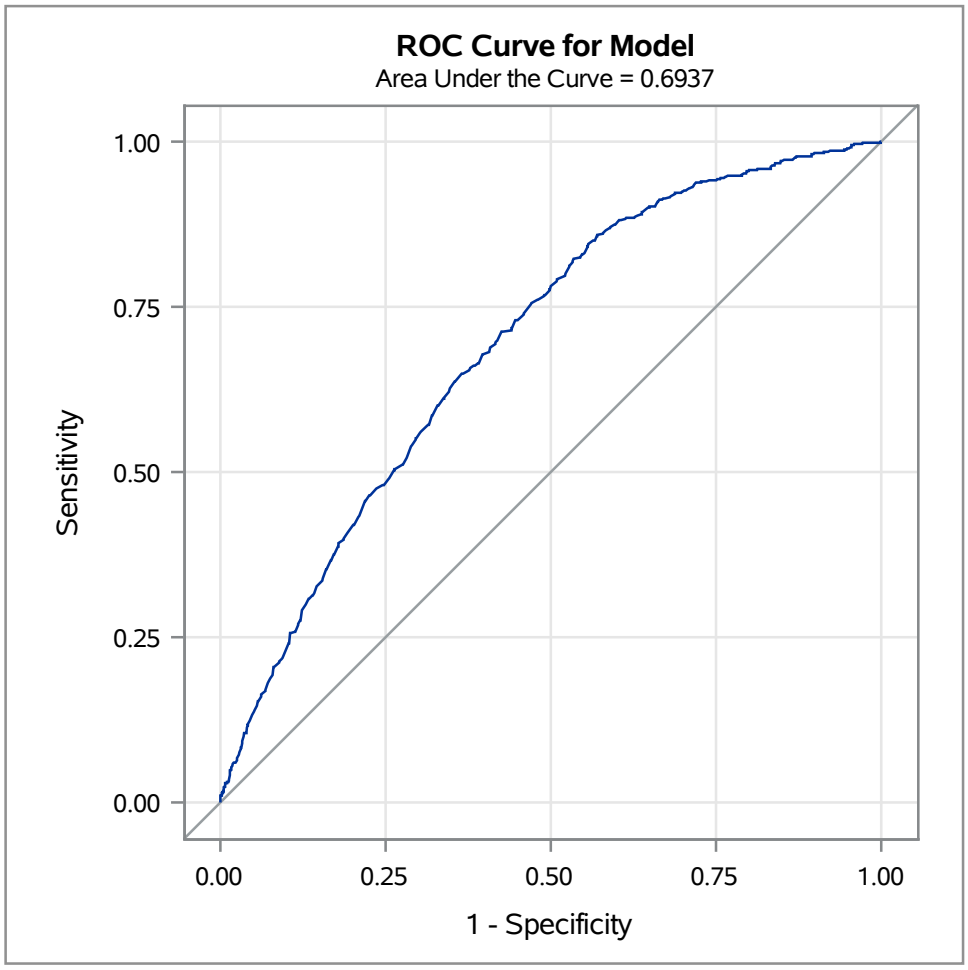
Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
LDL	1	112.4120	<.0001
leo	1	4.9378	0.0263
Gender	1	41.4670	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-3.4613	0.2268	232.8749	<.0001
LDL		1	0.0168	0.00158	112.4120	<.0001
leo	**0**	1	-0.1666	0.0750	4.9378	0.0263
Gender	**Female**	1	-0.5159	0.0801	41.4670	<.0001

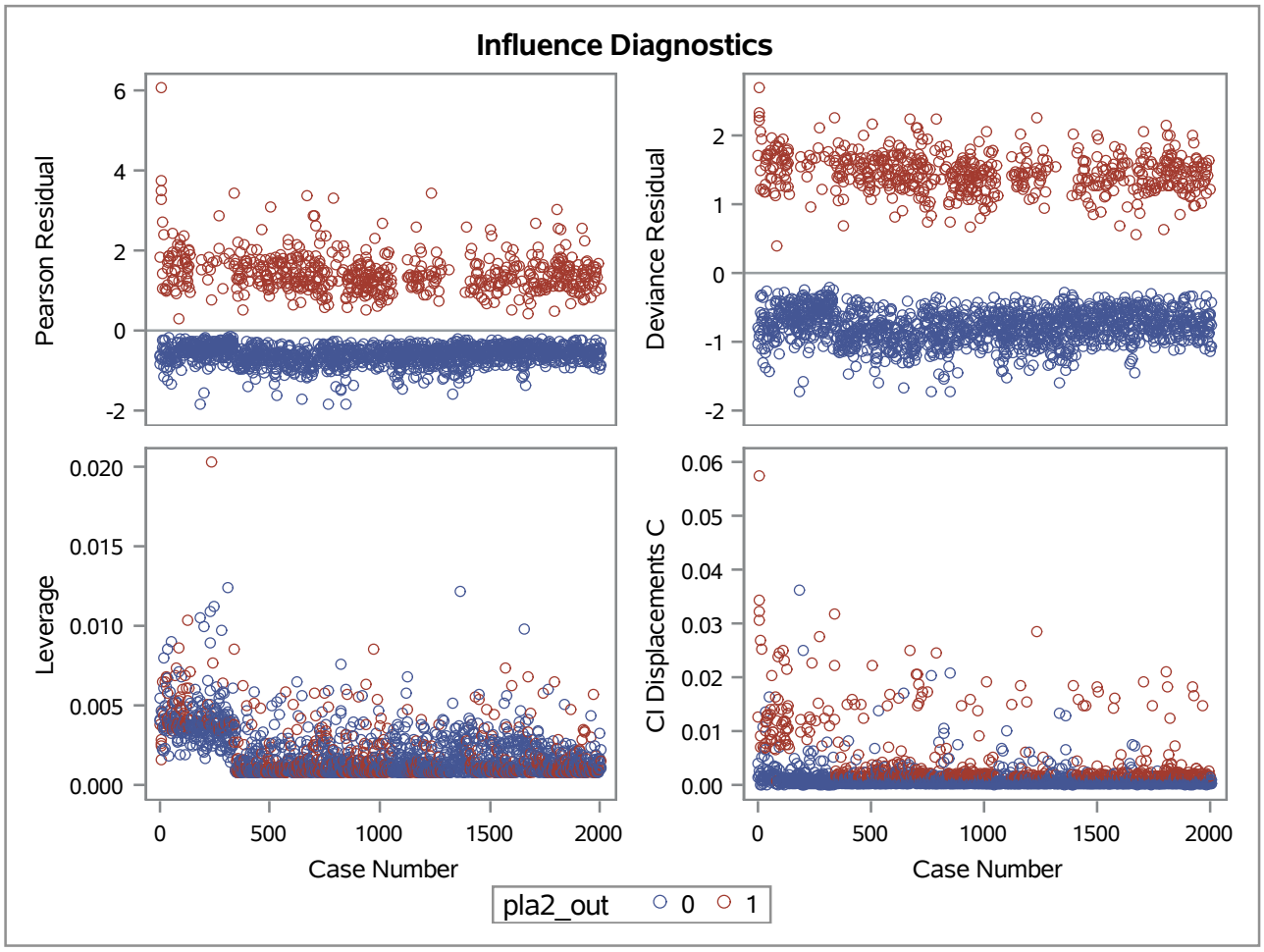
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
LDL	1.017	1.014	1.020
leo 0 vs 1	0.717	0.534	0.961
Gender Female vs Male	0.356	0.260	0.488

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.1	**Somers' D**	0.387
Percent Discordant	30.4	**Gamma**	0.389
Percent Tied	0.4	**Tau-a**	0.161
Pairs	809914	**c**	0.694

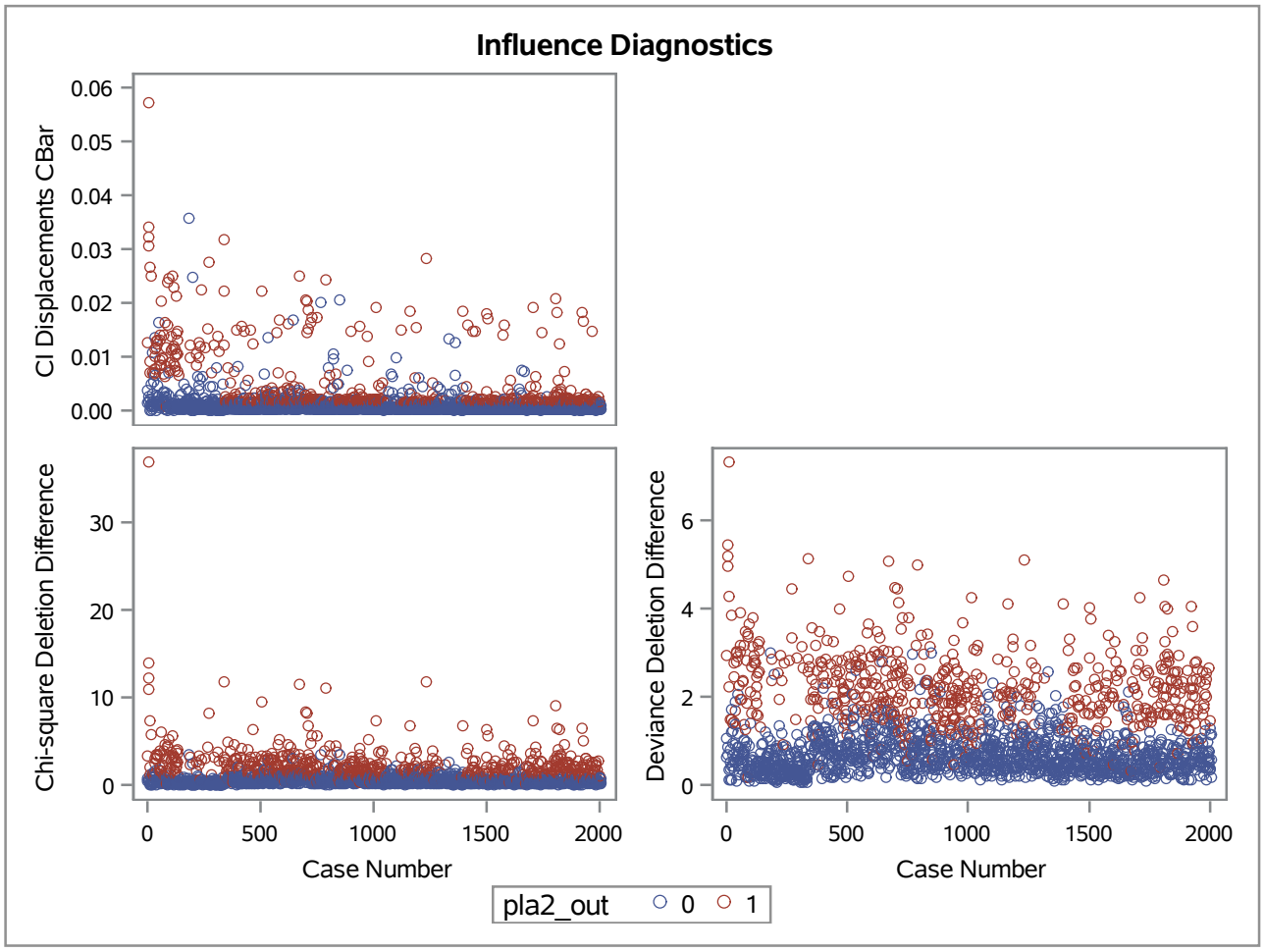
The LOGISTIC Procedure



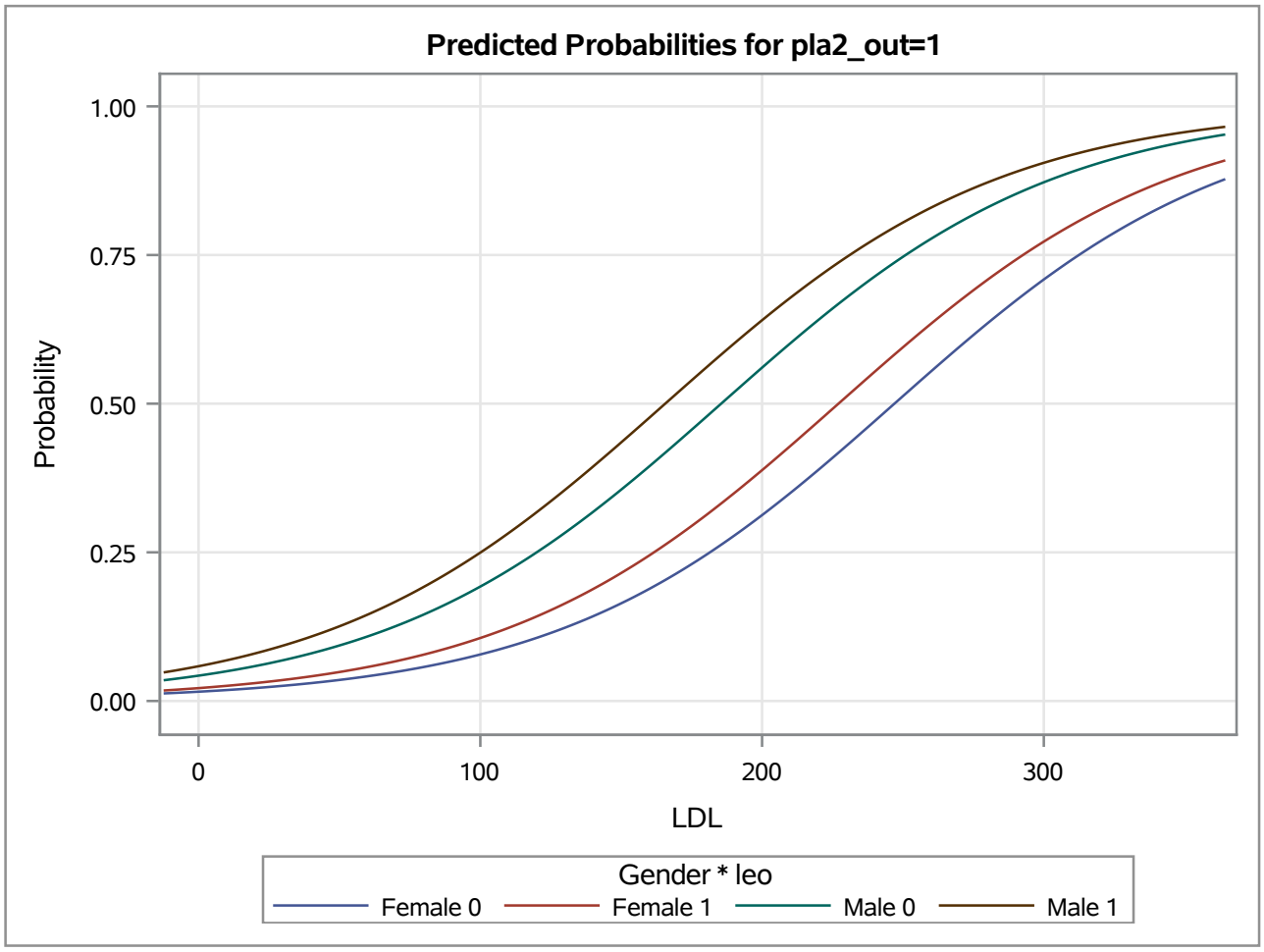
The LOGISTIC Procedure



The LOGISTIC Procedure



The LOGISTIC Procedure



The HPSPLIT Procedure

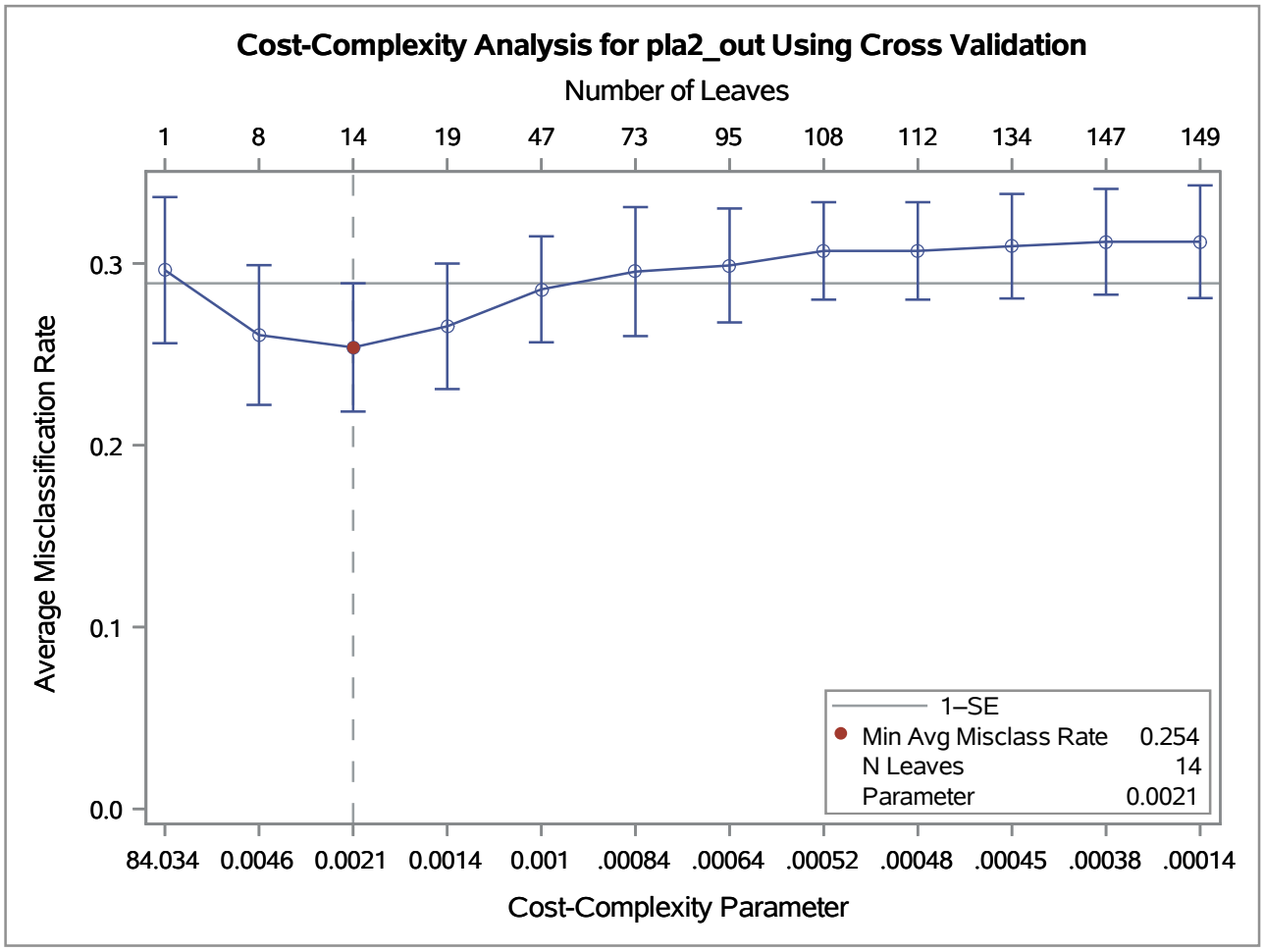
Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	**Engine**	**Role**	**Path**
WORK.TEMP1	V9	Input	On Client

Model Information	
Split Criterion Used	Entropy
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	10
Tree Depth	10
Number of Leaves Before Pruning	169
Number of Leaves After Pruning	14
Model Event Level	1

Number of Observations Read	2006
Number of Observations Used	2006

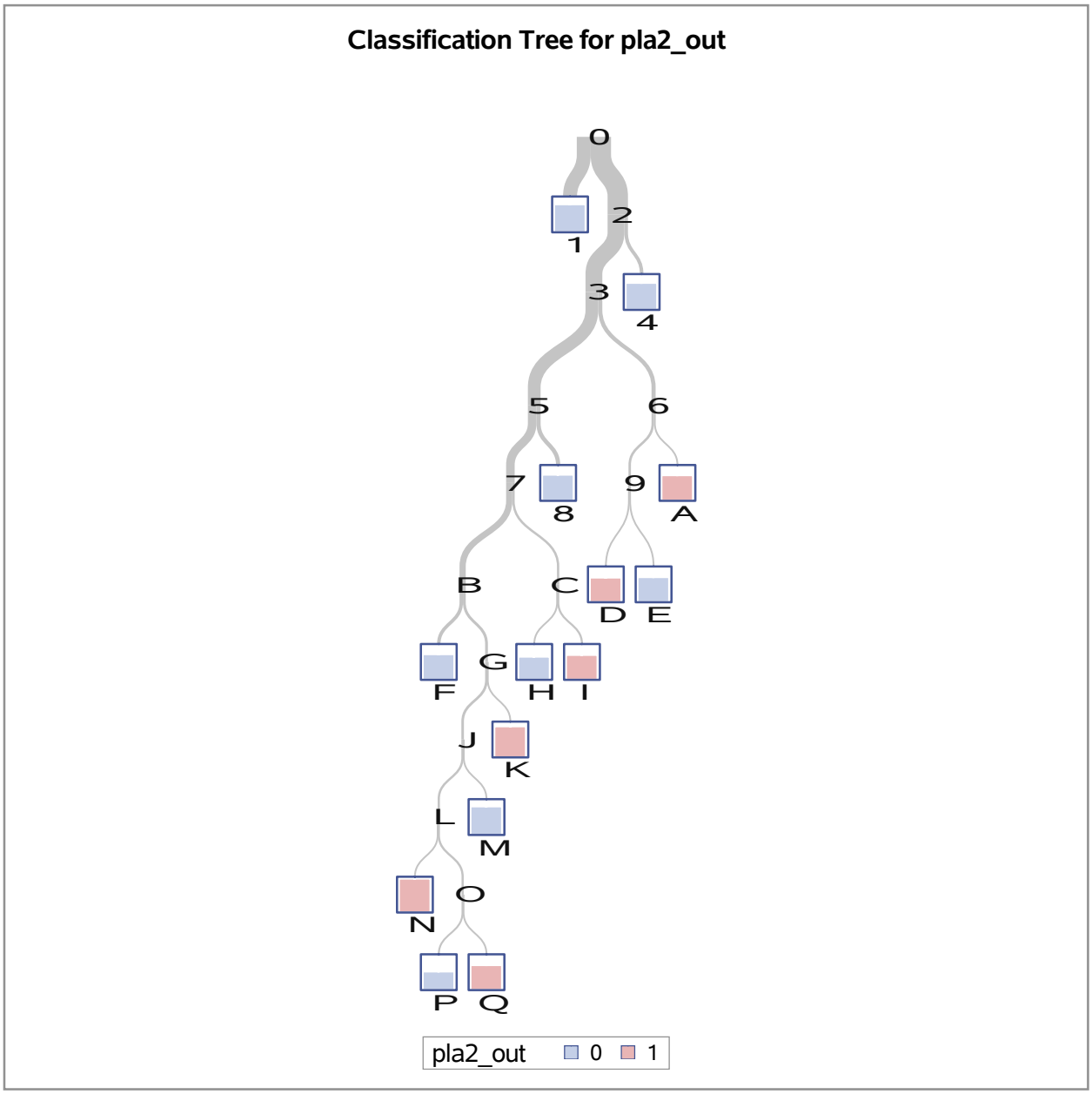
The HPSPLIT Procedure



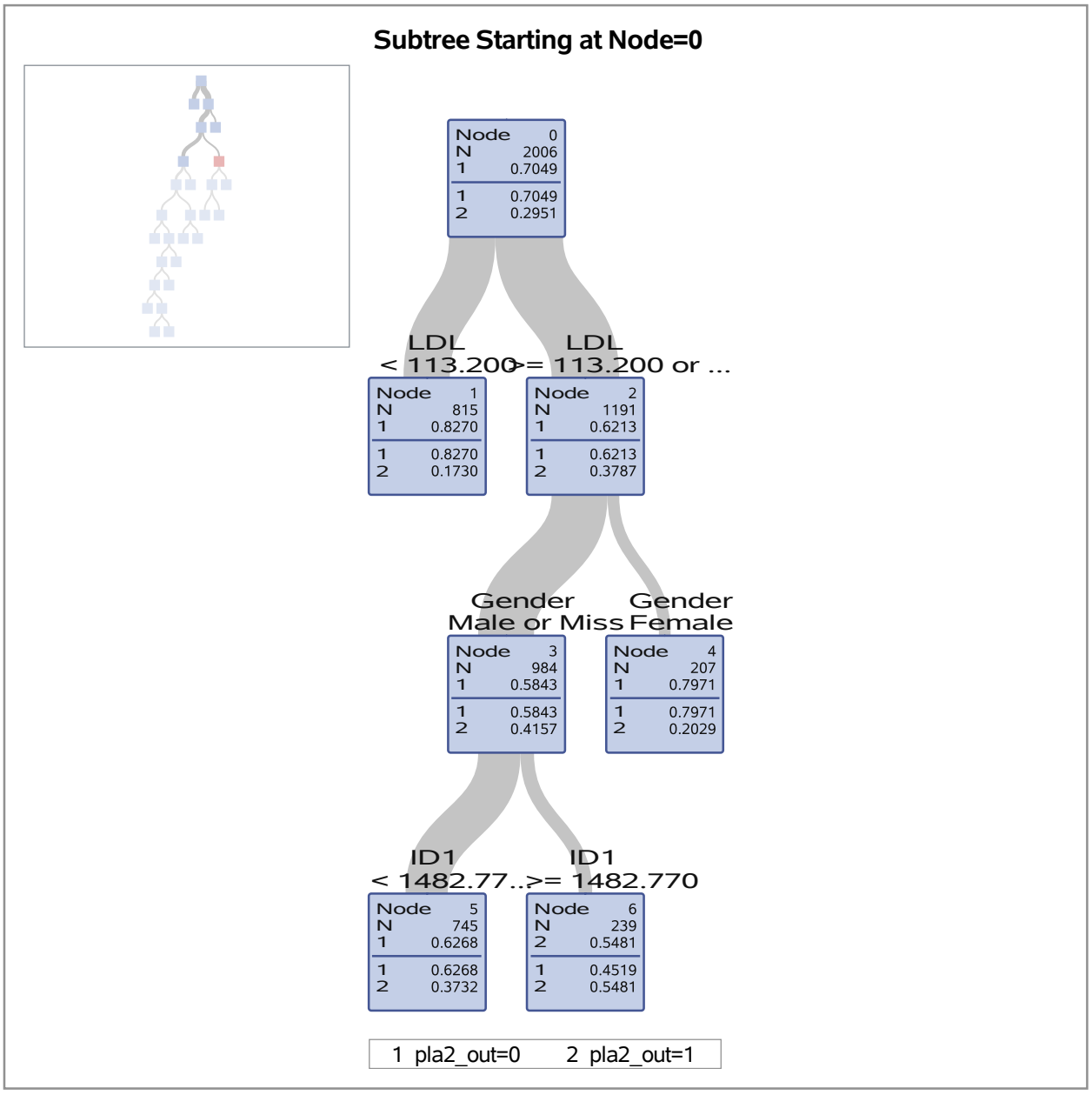
10-Fold Cross Validation Assessment of Model											
	Average Square Error				Number of Leaves			Misclassification Rate			
N Leaves	Min	Avg	Standard Error	Max	Min	Median	Max	Min	Avg	Standard Error	Max
10	0.1503	0.1834	0.0174	0.2184	10	10.5	15	0.1881	0.2530	0.0347	0.3131

10-Fold Cross Validation Confusion Matrix			
	Predicted		
Actual	0	1	Error Rate
0	1315	99	0.0700
1	408	184	0.6892

The HPSPLIT Procedure



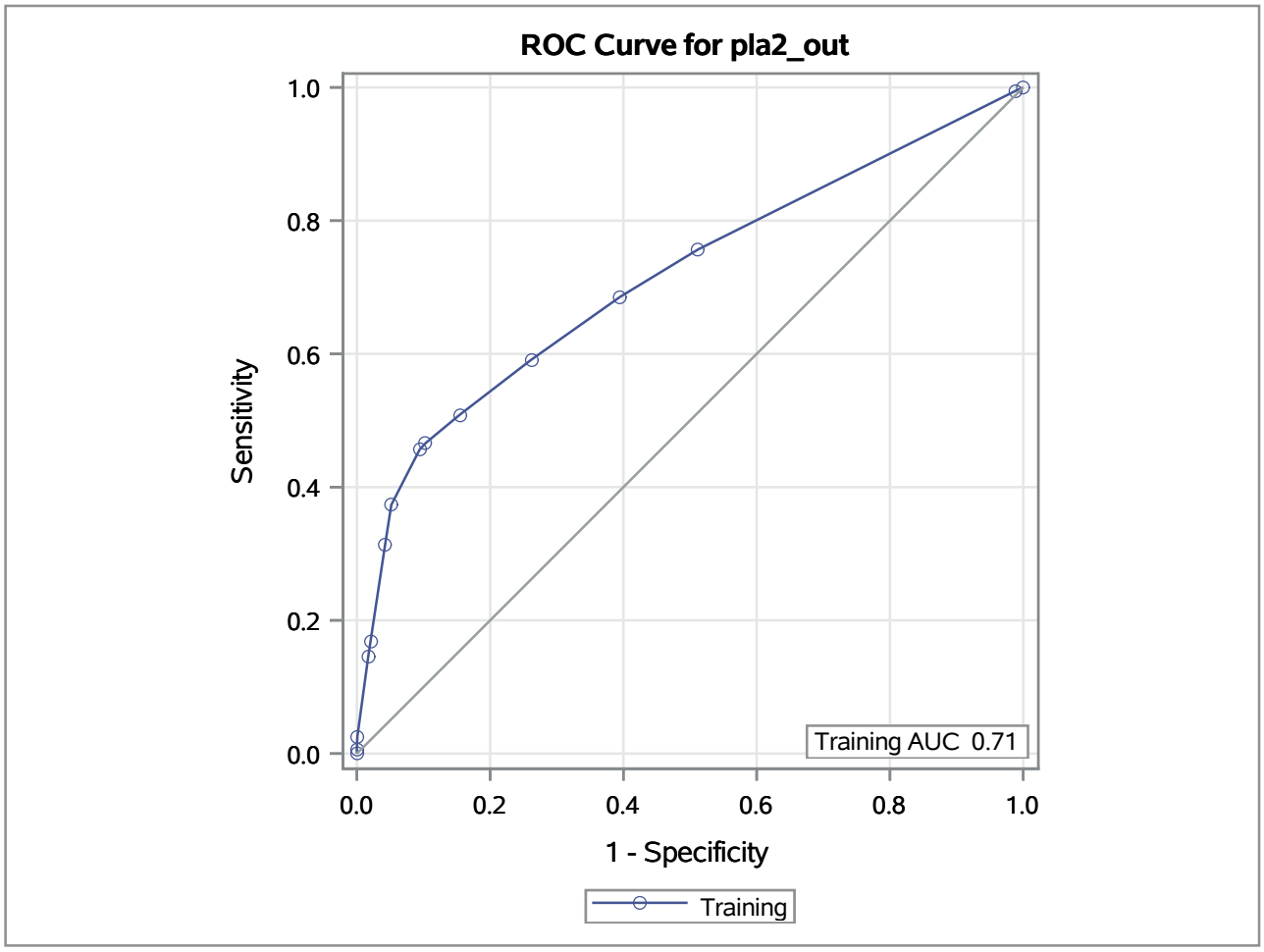
The HPSPLIT Procedure



The HPSPLIT Procedure

Confusion Matrices				
		Predicted		
	Actual	0	1	Error Rate
Model Based	0	1341	73	0.0516
	1	371	221	0.6267
Cross Validation	0	1315	99	0.0700
	1	408	184	0.6892

Fit Statistics for Selected Tree									
	N Leaves	ASE	Mis-class	Sensitivity	Specificity	Entropy	Gini	RSS	AUC
Model Based	14	0.1679	0.2213	0.3733	0.9484	0.7451	0.3357	673.5	0.7144
Cross Validation	10	0.1834	0.2530	0.3108	0.9300				



The HPSPLIT Procedure

Variable Importance				
		Training		
Variable	Variable Label	Relative	Importance	Count
LDL	LDL	1.0000	8.0343	4
ID1	ID1	0.6863	5.5138	3
HDL	HDL	0.6323	5.0803	2
IR	IR	0.5593	4.4935	1
Gender	Gender	0.4897	3.9347	1
wt		0.2034	1.6343	1
HR	HR	0.1760	1.4144	1

The HPFOREST Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	**Engine**	**Role**	**Path**
WORK.TEMP1	V9	Input	On Client
LEO.SCORE	V9	Output	On Client

Model Information		
Parameter	**Value**	
Variables to Try	4	(Default)
Maximum Trees	25	
Actual Trees	25	
Inbag Fraction	0.3	
Prune Fraction	0	(Default)
Prune Threshold	0.1	(Default)
Leaf Fraction	0.00001	(Default)
Leaf Size Setting	1	(Default)
Leaf Size Used	1	
Category Bins	30	(Default)
Interval Bins	100	
Minimum Category Size	5	(Default)
Node Size	100000	(Default)
Maximum Depth	20	(Default)
Alpha	1	(Default)
Exhaustive	5000	(Default)
Rows of Sequence to Skip	5	(Default)
Split Criterion	.	Gini
Preselection Method	.	BinnedSearch
Missing Value Handling	.	Valid value

Number of Observations	
Type	**N**
Number of Observations Read	2006
Number of Observations Used	2006

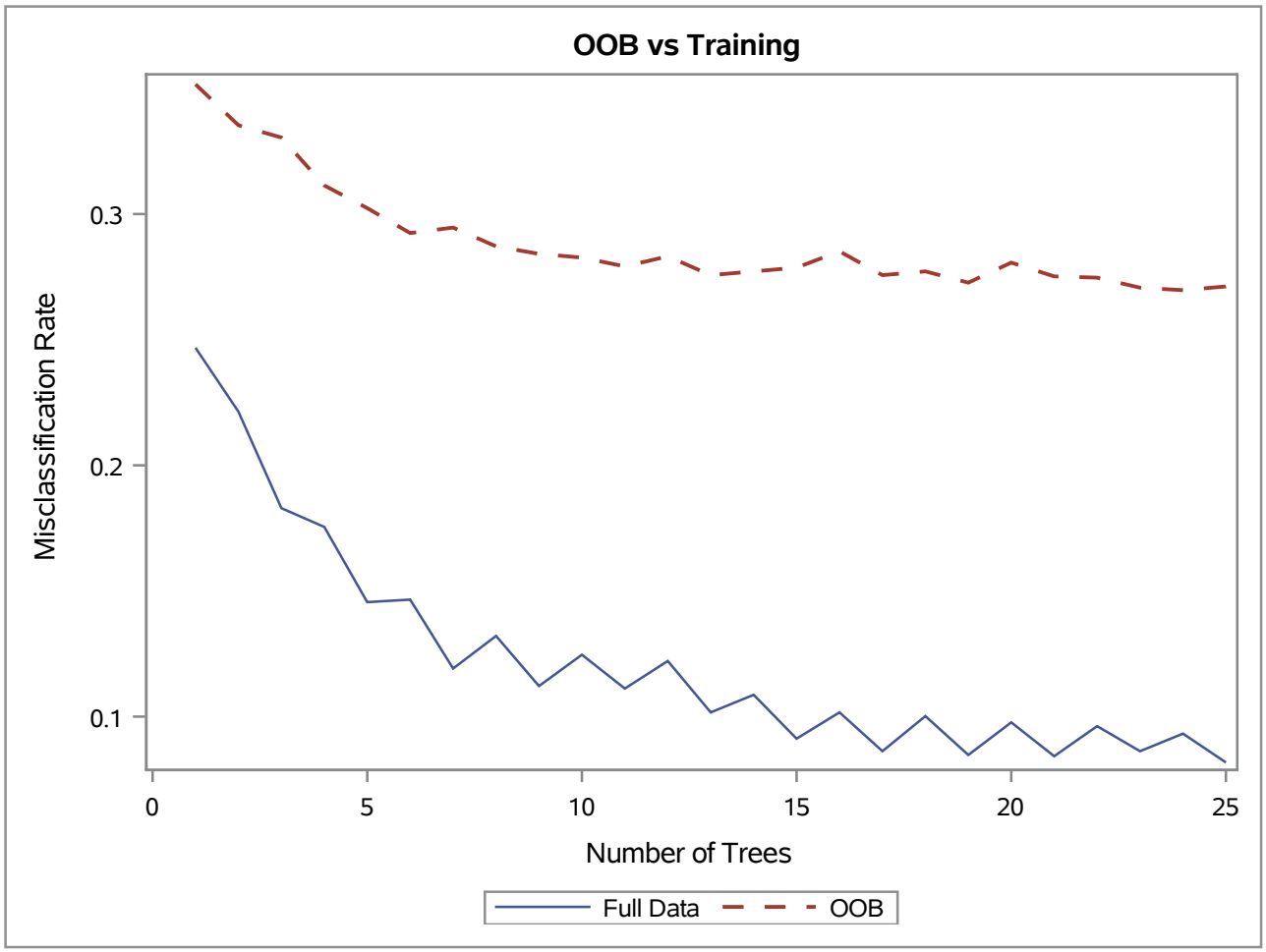
The HPFOREST Procedure

Baseline Fit Statistics	
Statistic	Value
Average Square Error	0.208
Misclassification Rate	0.295
Log Loss	0.607

Fit Statistics							
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)	Misclassification Rate (Train)	Misclassification Rate (OOB)	Log Loss (Train)	Log Loss (OOB)
1	140	0.2466	0.352	0.2468	0.352	5.651	8.068
2	264	0.1653	0.318	0.2213	0.335	1.953	6.231
3	394	0.1364	0.275	0.1830	0.330	0.927	4.448
4	528	0.1201	0.242	0.1755	0.311	0.541	3.141
5	666	0.1114	0.224	0.1456	0.302	0.415	2.297
6	795	0.1045	0.211	0.1466	0.292	0.373	1.798
7	929	0.1015	0.204	0.1191	0.295	0.348	1.499
8	1054	0.0989	0.198	0.1321	0.287	0.324	1.253
9	1172	0.0970	0.192	0.1122	0.284	0.320	1.027
10	1300	0.0958	0.190	0.1246	0.283	0.319	0.902
11	1434	0.0951	0.189	0.1112	0.279	0.319	0.880
12	1557	0.0938	0.187	0.1221	0.283	0.317	0.804
13	1693	0.0931	0.185	0.1017	0.276	0.316	0.730
14	1831	0.0913	0.182	0.1087	0.277	0.313	0.683
15	1957	0.0902	0.180	0.0912	0.279	0.310	0.647
16	2086	0.0894	0.179	0.1017	0.285	0.310	0.634
17	2215	0.0889	0.179	0.0862	0.276	0.309	0.624
18	2346	0.0891	0.179	0.1002	0.277	0.310	0.625
19	2482	0.0881	0.178	0.0847	0.273	0.308	0.613
20	2621	0.0875	0.177	0.0977	0.281	0.307	0.602
21	2737	0.0874	0.177	0.0842	0.275	0.307	0.600
22	2862	0.0871	0.176	0.0962	0.275	0.307	0.580
23	2989	0.0870	0.176	0.0862	0.271	0.308	0.571
24	3115	0.0869	0.176	0.0932	0.270	0.308	0.560
25	3241	0.0867	0.175	0.0818	0.271	0.307	0.559

The HPFOREST Procedure

Loss Reduction Variable Importance					
Variable	Number of Rules	Gini	OOB Gini	Margin	OOB Margin
Gender	14	0.004934	0.00359	0.009868	0.00819
leo	9	0.001437	-0.00074	0.002873	0.00091
IR	187	0.029235	-0.01032	0.058469	0.01819
LDL	233	0.037366	-0.01088	0.074732	0.02687
HbA1c	155	0.018902	-0.01327	0.037804	0.00478
ApoB	160	0.024672	-0.01423	0.049344	0.00964
HDL	190	0.026136	-0.01456	0.052272	0.01129
ID1	272	0.042531	-0.01688	0.085063	0.02422
AGE	136	0.017580	-0.01883	0.035159	-0.00050
HR	158	0.018781	-0.01913	0.037562	0.00022
BMI	163	0.020322	-0.01936	0.040645	0.00123
Insulin	186	0.019294	-0.02127	0.038589	-0.00263
sdLDL_C	291	0.033908	-0.02145	0.067816	0.01532
hc_CRP	250	0.027320	-0.02223	0.054639	0.00529
TC	267	0.036552	-0.02252	0.073104	0.01420
TG	249	0.027225	-0.02667	0.054451	0.00010
wt	296	0.030363	-0.02929	0.060726	0.00111


OOB vs Training
Misclassification Rate
0.3
0.2
0.1
0
5
10
15
20
25
Number of Trees
Full Data
OOB

The FREQ Procedure

Frequency Col Pct	Table of pla2_out by pred		
	pred		
pla2_out	0	1	Total
0	1397 90.48	17 3.68	1414
1	147 9.52	445 96.32	592
Total	1544	462	2006

Statistics for Table of pla2_out by pred

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	248.7655	148.8896	415.6387
Relative Risk (Column 1)	3.9788	3.4581	4.5780
Relative Risk (Column 2)	0.0160	0.0099	0.0257

Sample Size = 2006

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEMP1
Response Variable	pla2_out
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	2006
Number of Observations Used	961

Response Profile		
Ordered Value	**pla2_out**	**Total Frequency**
1	0	620
2	1	341

Probability modeled is pla2_out=1.

Note: 1045 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information		
Class	**Value**	**Design Variables**
Gender	**Female**	1
	Male	-1
leo	**0**	1
	1	-1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1252.051	1112.019
SC	1256.919	1136.359
-2 Log L	1250.051	1102.019

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	148.0315	4	<.0001
Score	136.8556	4	<.0001
Wald	114.8898	4	<.0001

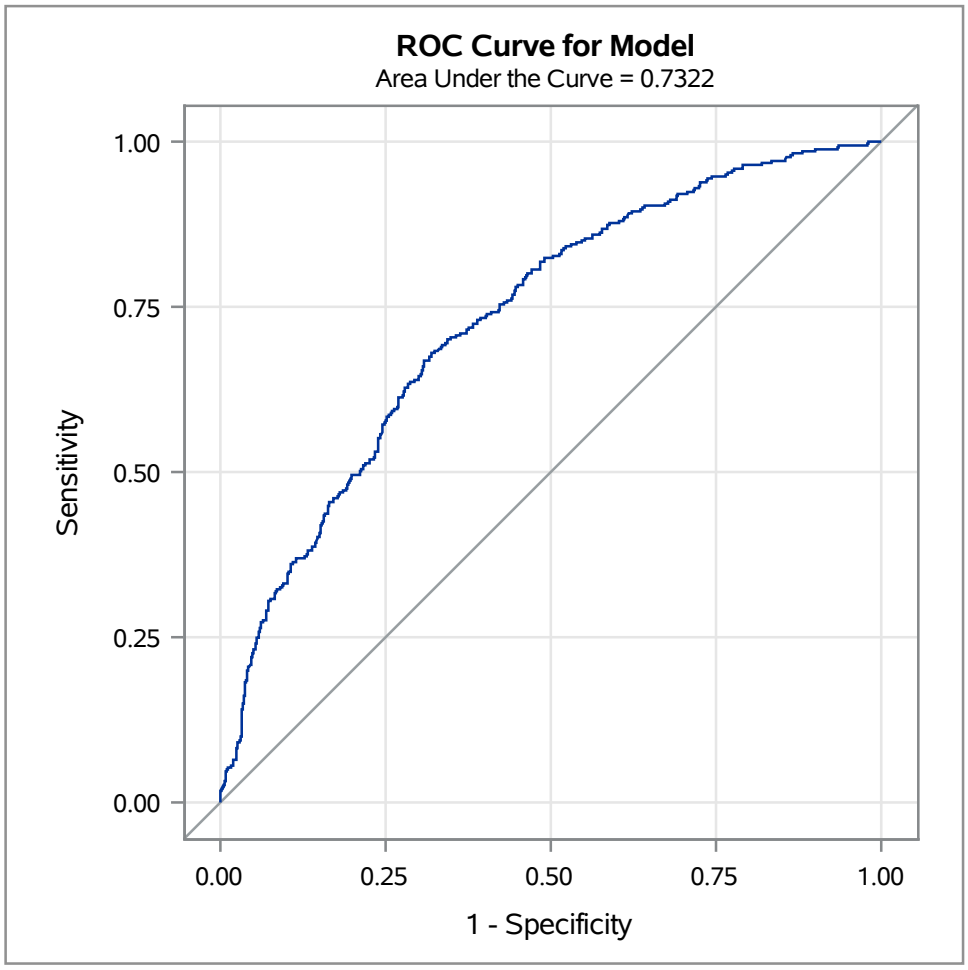
Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
LDL	1	94.3640	<.0001
leo	1	1.0348	0.3090
Gender	1	21.2254	<.0001
IR	1	7.2165	0.0072

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-4.0589	0.3552	130.5925	<.0001
LDL		1	0.0232	0.00238	94.3640	<.0001
leo	**0**	1	0.1082	0.1063	1.0348	0.3090
Gender	**Female**	1	-0.5188	0.1126	21.2254	<.0001
IR		1	0.00789	0.00294	7.2165	0.0072

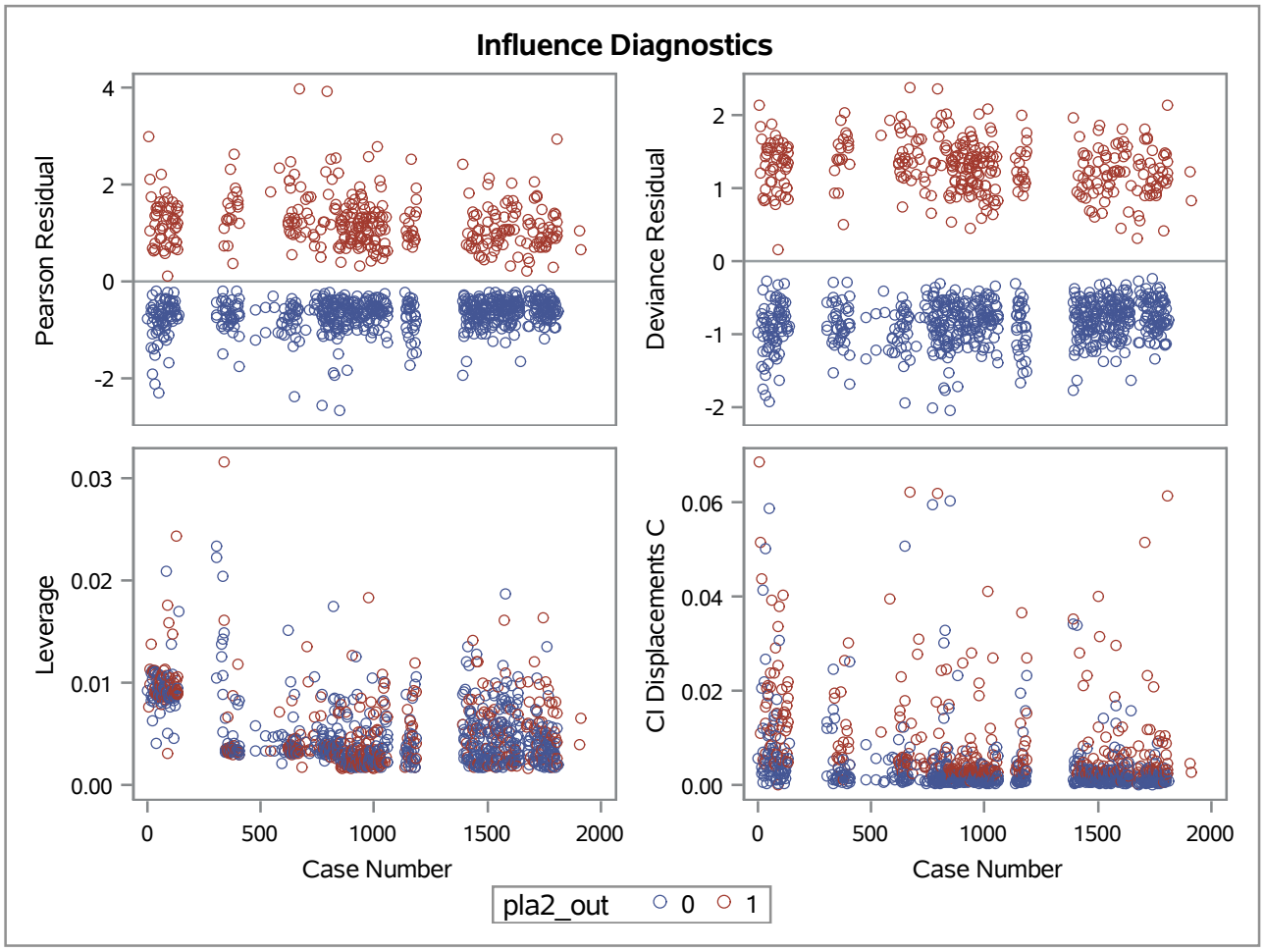
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
LDL	1.023	1.019	1.028
leo 0 vs 1	1.241	0.818	1.883
Gender Female vs Male	0.354	0.228	0.551
IR	1.008	1.002	1.014

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	73.2	**Somers' D**	0.464
Percent Discordant	26.8	**Gamma**	0.464
Percent Tied	0.0	**Tau-a**	0.213
Pairs	211420	**c**	0.732

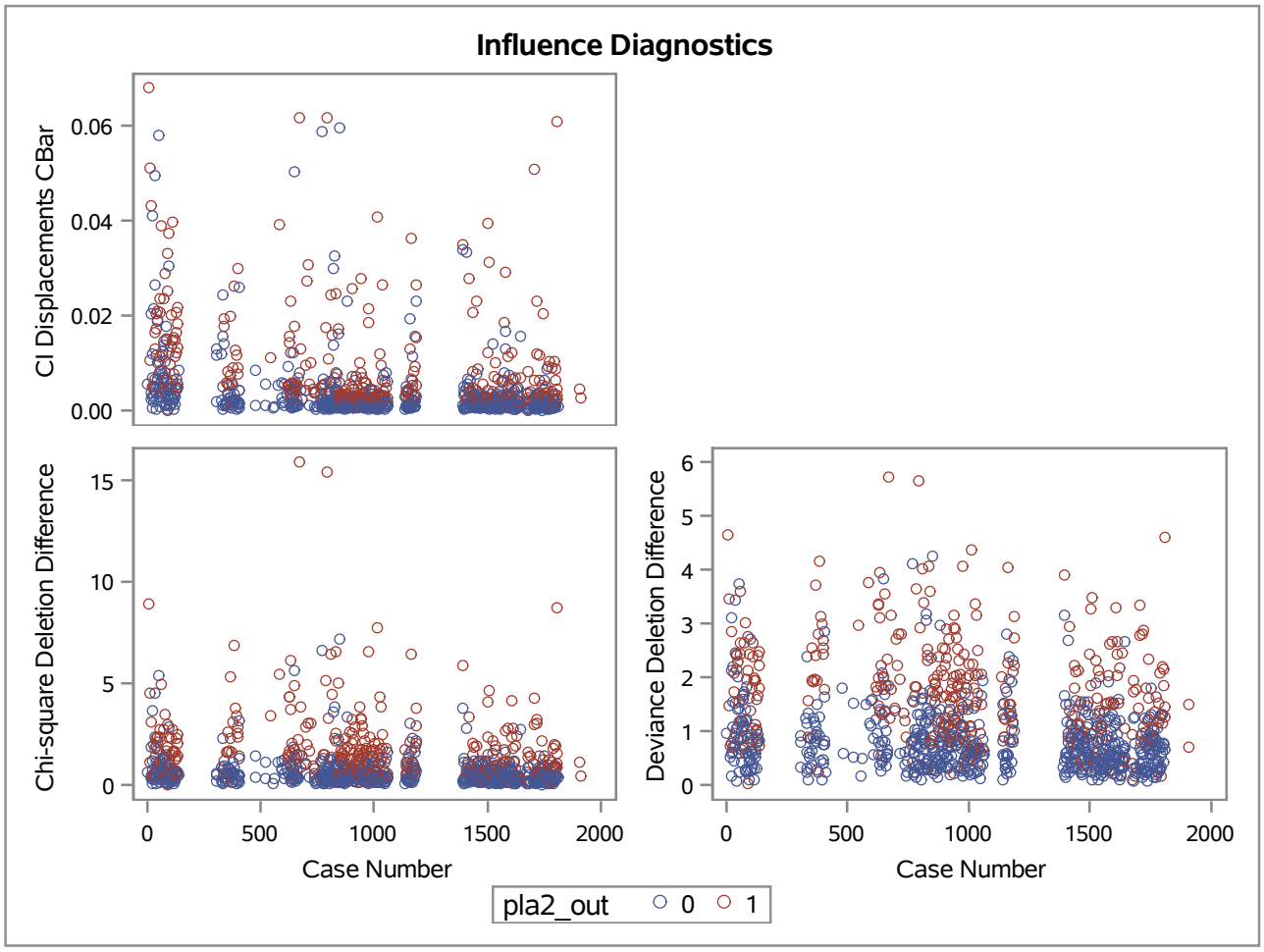
The LOGISTIC Procedure



The LOGISTIC Procedure



The LOGISTIC Procedure



The LOGISTIC Procedure