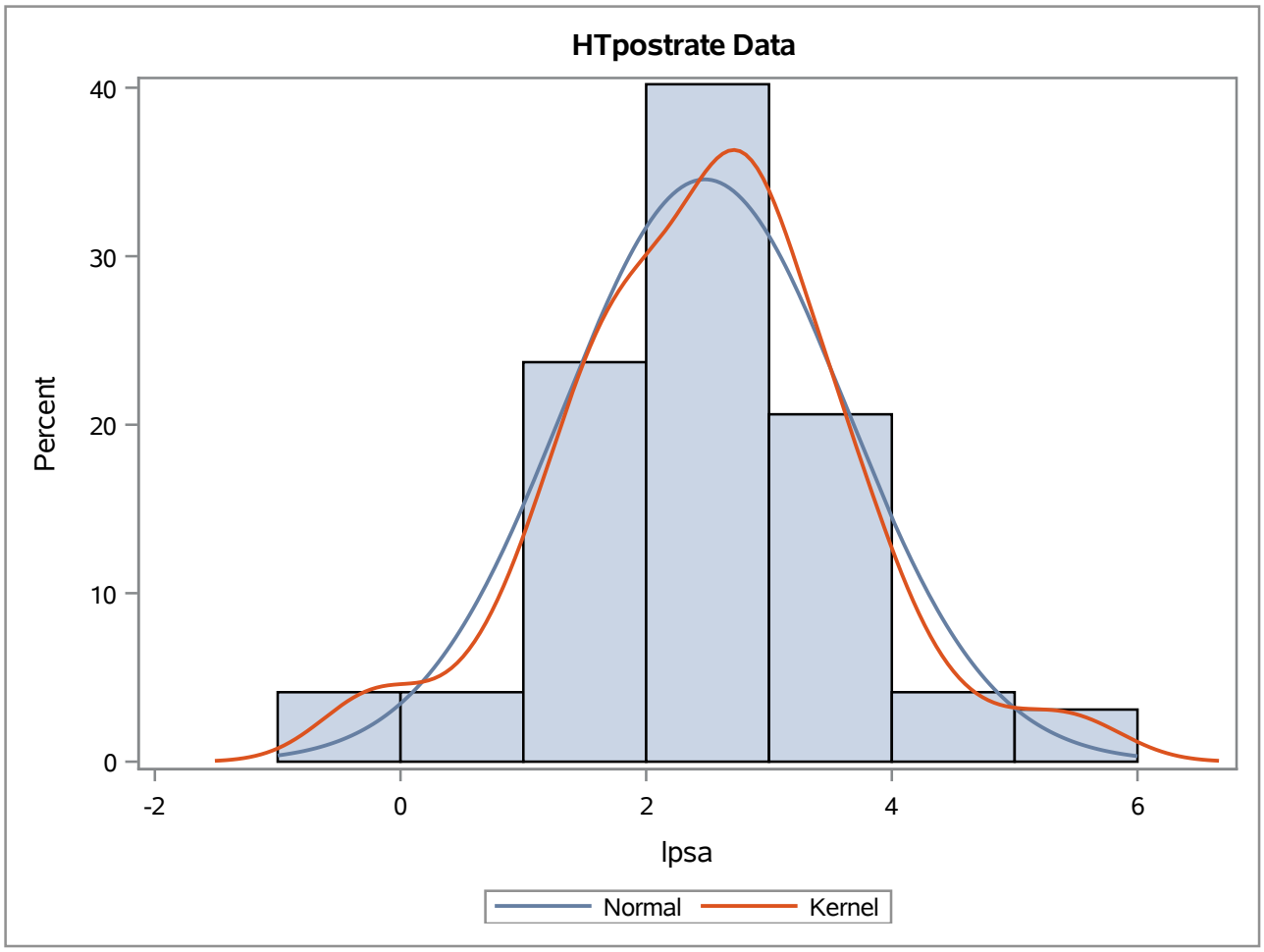
The MEANS Procedure

Variable	Label	Lower Quartile	Median	Upper Quartile
age	age	60.0000000	65.0000000	68.0000000
gleason	gleason	6.0000000	7.0000000	7.0000000
lbph	lbph	-1.3862944	0.3001046	1.5581446
lcavol	lcavol	0.5128236	1.4469190	2.1270405
lcp	lcp	-1.3862944	-0.7985077	1.1786550
lweight	lweight	3.3758800	3.6230070	3.8763960
pgg45	pgg45	0	15.0000000	40.0000000


HTpostrate Data
Percent
40
30
20
10
0
-2
0
2
4
6
lpsa
Normal
Kernel

The GLMSELECT Procedure

Data Set	WORK.PROSTRATE
Dependent Variable	lpsa
Selection Method	Stepwise
Select Criterion	Significance Level
Stop Criterion	PRESS
Entry Significance Level (SLE)	0.15
Stay Significance Level (SLS)	0.15
Effect Hierarchy Enforced	None

Number of Observations Read	67
Number of Observations Used	67

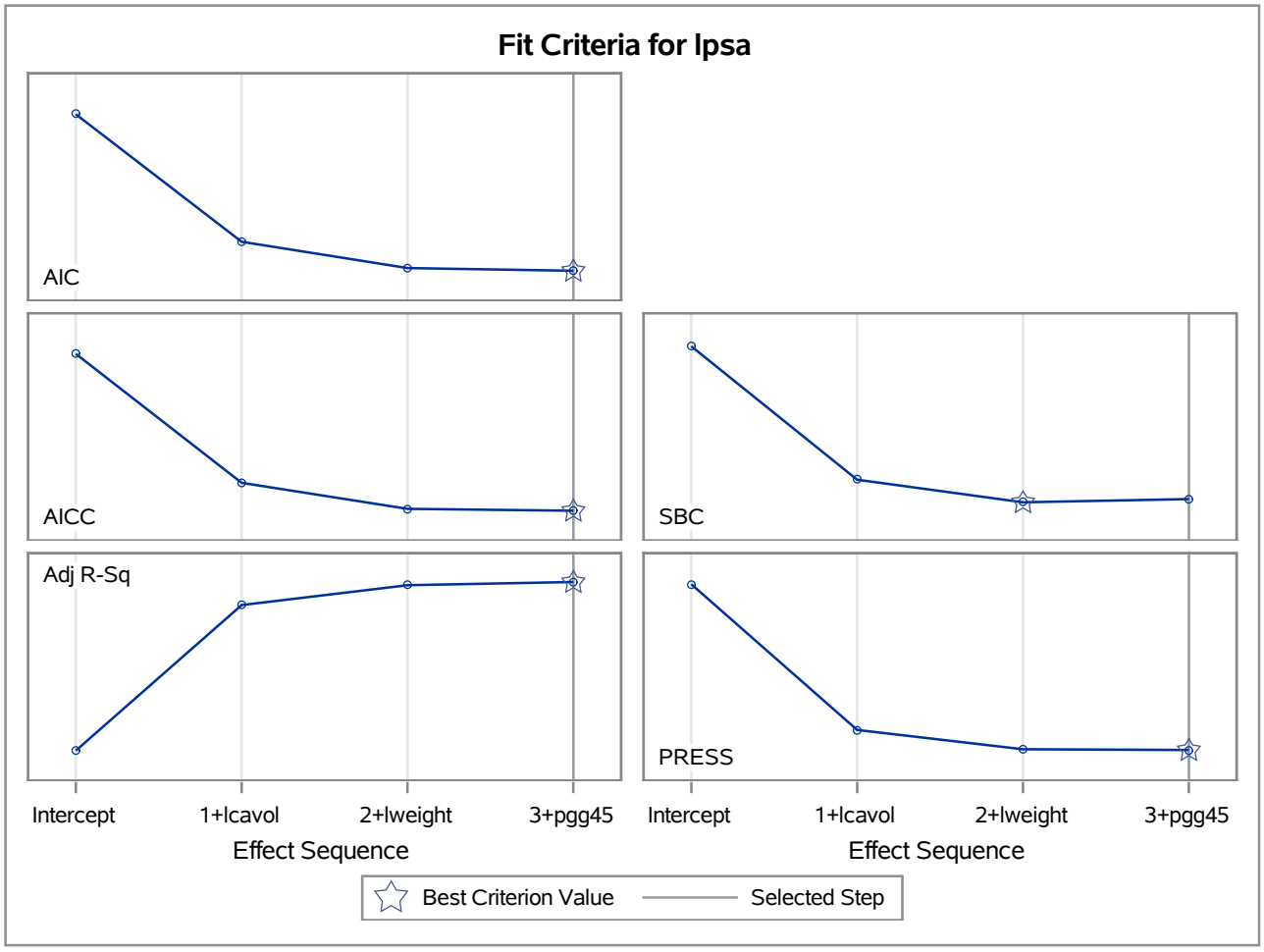
Dimensions	
Number of Effects	8
Number of Parameters	8

The GLMSELECT Procedure

Stepwise Selection Summary						
Step	Effect Entered	Effect Removed	Number Effects In	PRESS	F Value	Pr > F
0	Intercept		1	99.2212	0.00	1.0000
1	lcavol		2	47.3773	75.55	<.0001
2	lweight		3	40.5033	12.83	0.0007
3	pgg45		4	40.2461*	2.95	0.0909
* Optimal Value of Criterion						

Selection stopped at a local minimum of the PRESS criterion.

Stop Details				
Candidate For	Effect	Candidate PRESS		Compare PRESS
Entry	lbph	40.5011	>	40.2461
Removal	pgg45	40.5033	>	40.2461



The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 3).

Effects:	Intercept lcavol lweight pgg45

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	3	60.84741	20.28247	36.06
Error	63	35.43403	0.56244	
Corrected Total	66	96.28145		

Root MSE	0.74996
Dependent Mean	2.45235
R-Square	0.6320
Adj R-Sq	0.6144
AIC	34.31967
AICC	35.30328
PRESS	40.24608
SBC	-25.86156

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-1.222729	0.725248	-1.69
lcavol	1	0.553546	0.089034	6.22
lweight	1	0.768076	0.203796	3.77
pgg45	1	0.006200	0.003611	1.72

The HPSPLIT Procedure

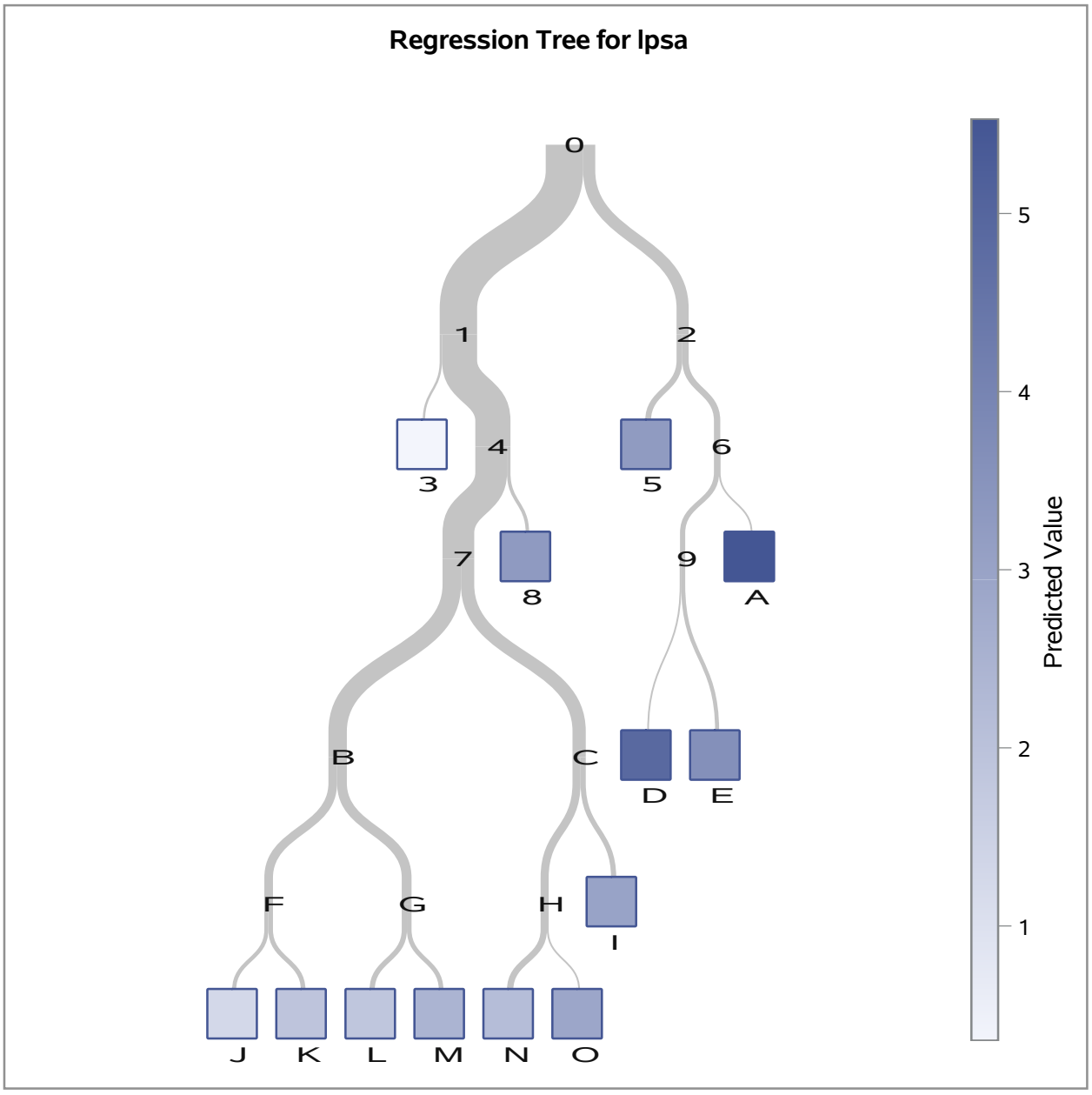
Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	**Engine**	**Role**	**Path**
WORK.PROSTRATE	V9	Input	On Client
WORK.HPSPLOUT	V9	Output	On Client

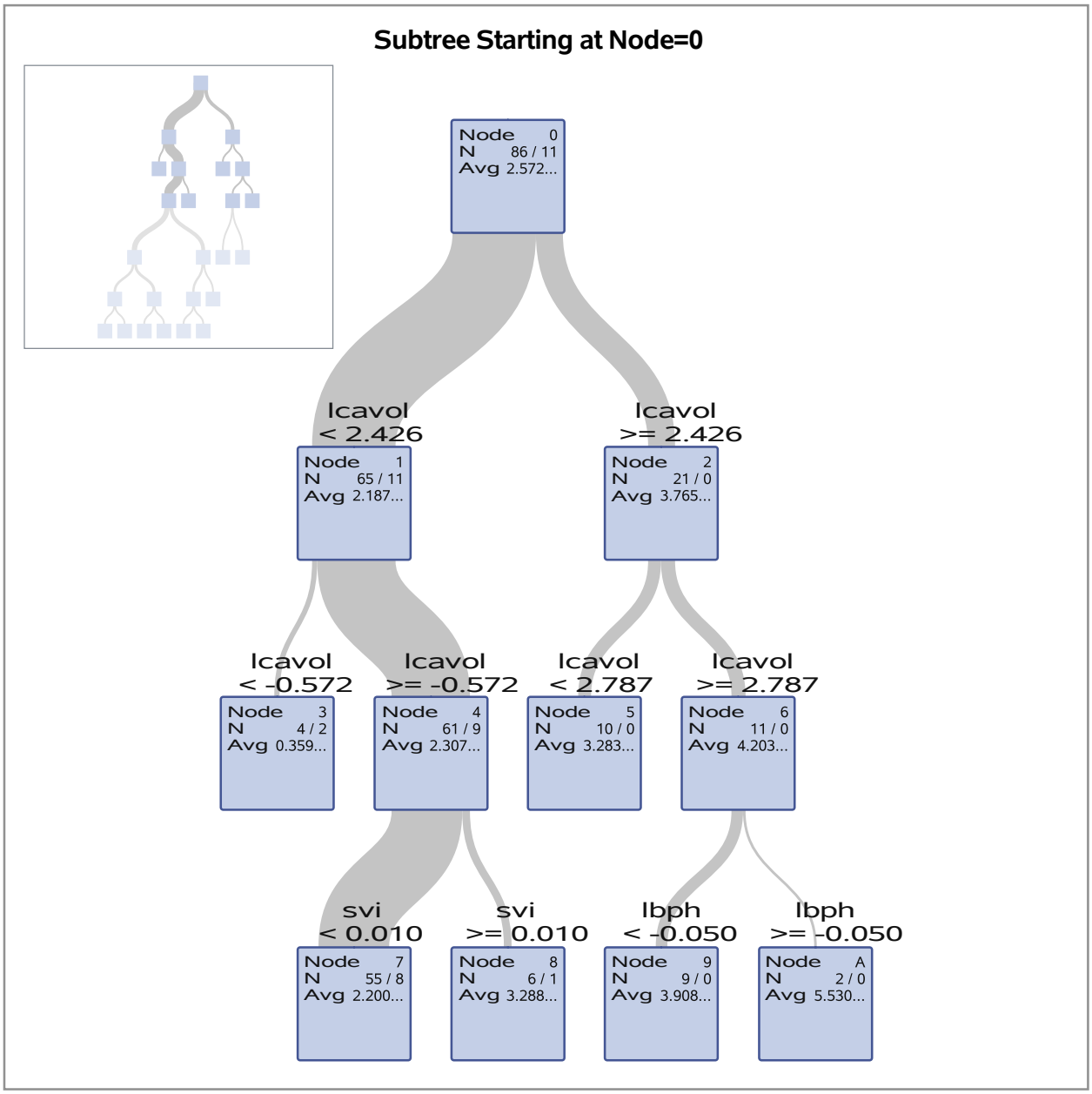
Model Information	
Split Criterion Used	Variance
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	10
Tree Depth	6
Number of Leaves Before Pruning	80
Number of Leaves After Pruning	13

Number of Observations Read	97
Number of Observations Used	97
Number of Training Observations Used	86
Number of Validation Observations Used	11

The HPSPLIT Procedure



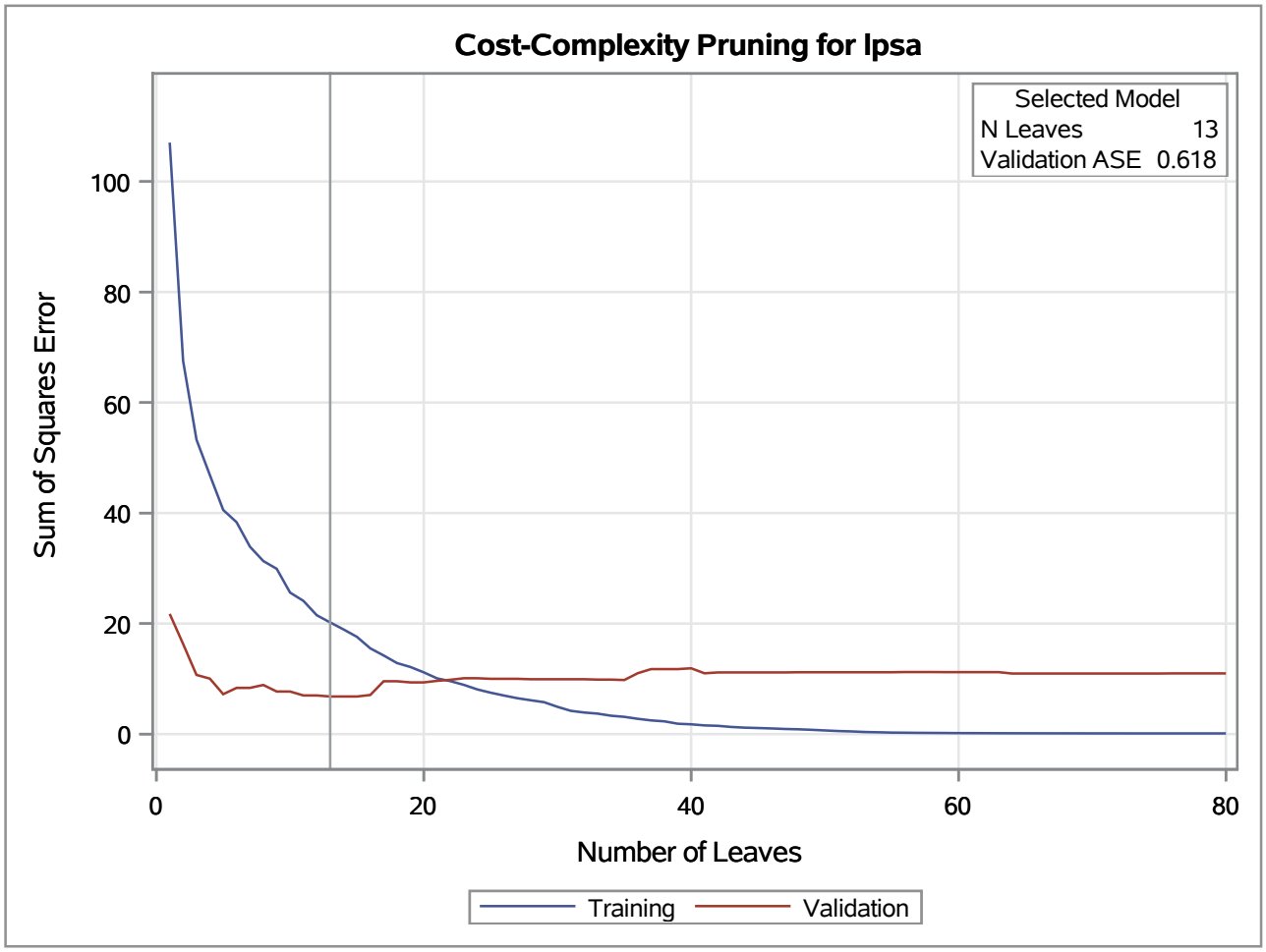
The HPSPLIT Procedure

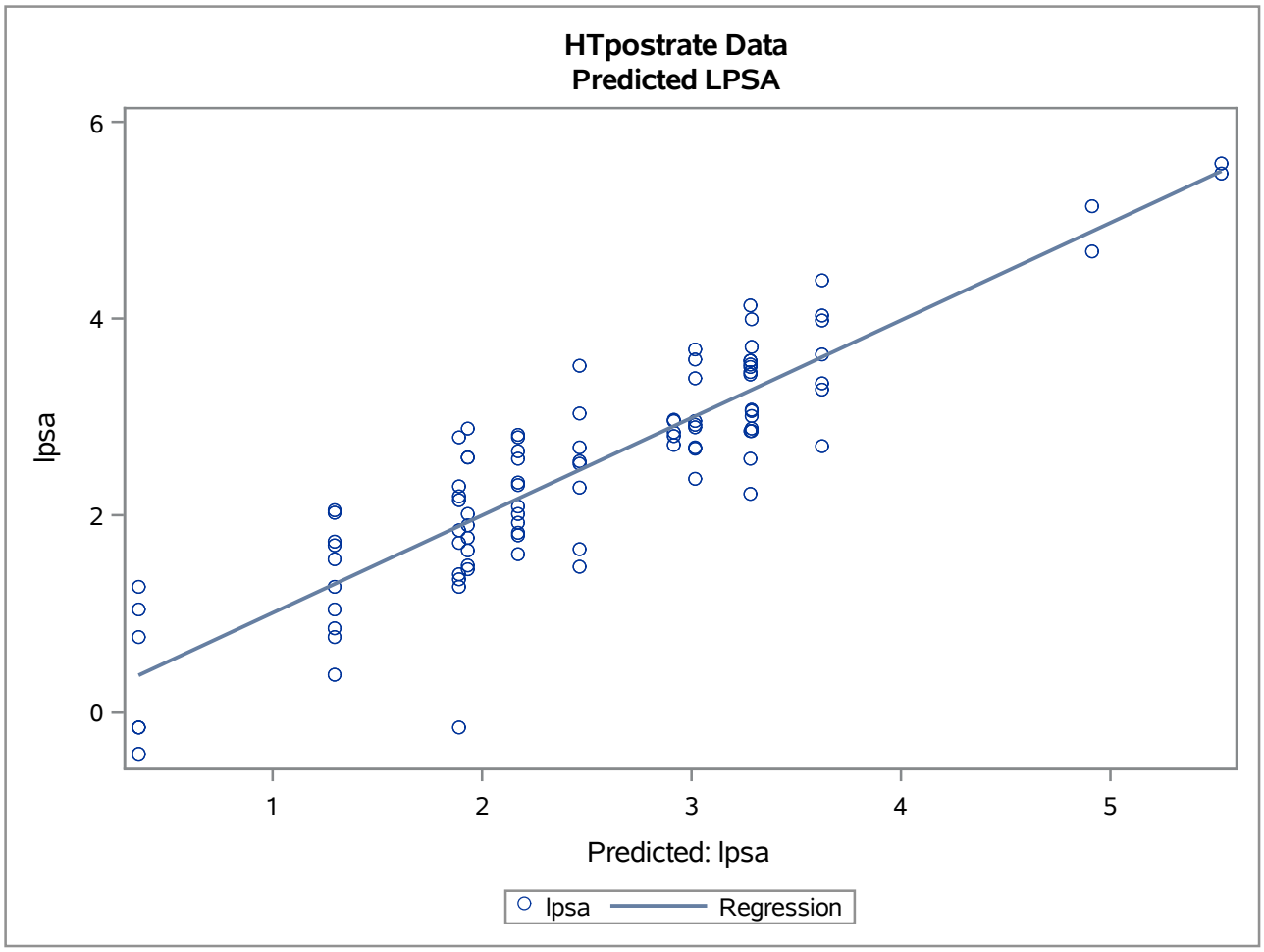


The HPSPLIT Procedure

Fit Statistics for Selected Tree			
	N Leaves	ASE	RSS
Training	13	0.2350	20.2083
Validation	13	0.6184	6.8024

Variable Importance							
		Training		Validation			
Variable	Variable Label	Relative	Importance	Relative	Importance	Relative Ratio	Count
lcavol	lcavol	1.0000	7.8920	1.0000	3.4278	1.0000	5
lweight	lweight	0.3497	2.7599	0.5054	1.7324	1.4452	2
lbph	lbph	0.3026	2.3881	0.3174	1.0879	1.0488	2
svi	svi	0.3207	2.5312	0.2395	0.8209	0.7467	1
age	age	0.2043	1.6122	0.0000	0	0.0000	1
gleason	gleason	0.1888	1.4901	0.0000	0	0.0000	1




HTpostrate Data
Predicted LPSA
lpsa
6
4
2
0
1
2
3
4
5
Predicted: lpsa
lpsa
Regression

The HPFOREST Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

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

Model Information		
Parameter	**Value**	
Variables to Try	3	(Default)
Maximum Trees	40	
Actual Trees	40	
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	.	Variance
Preselection Method	.	BinnedSearch
Missing Value Handling	.	Valid value

Number of Observations	
Type	**N**
Number of Observations Read	97
Number of Observations Used	97

The HPFOREST Procedure

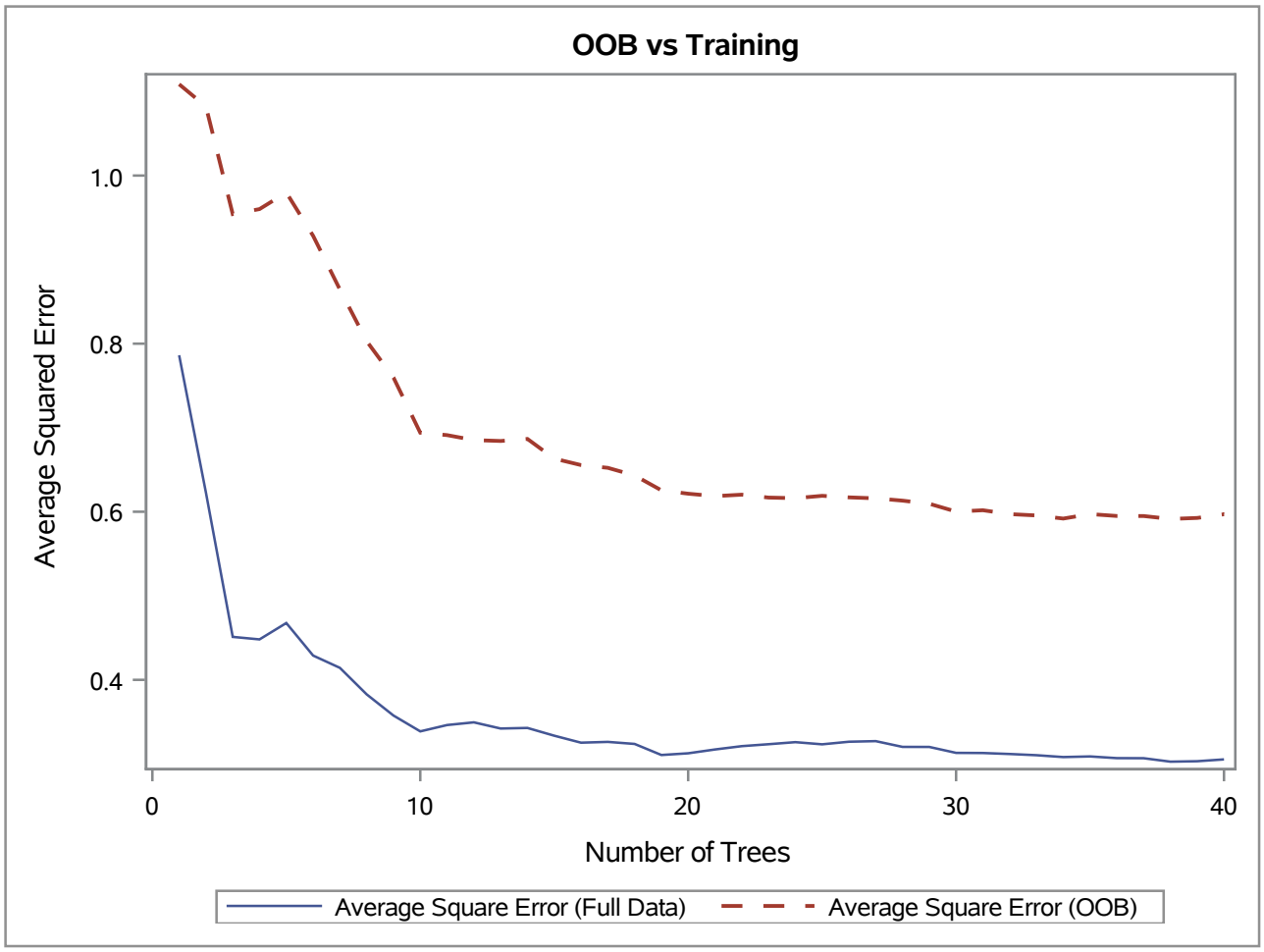
Baseline Fit Statistics	
Statistic	Value
Average Square Error	1.319

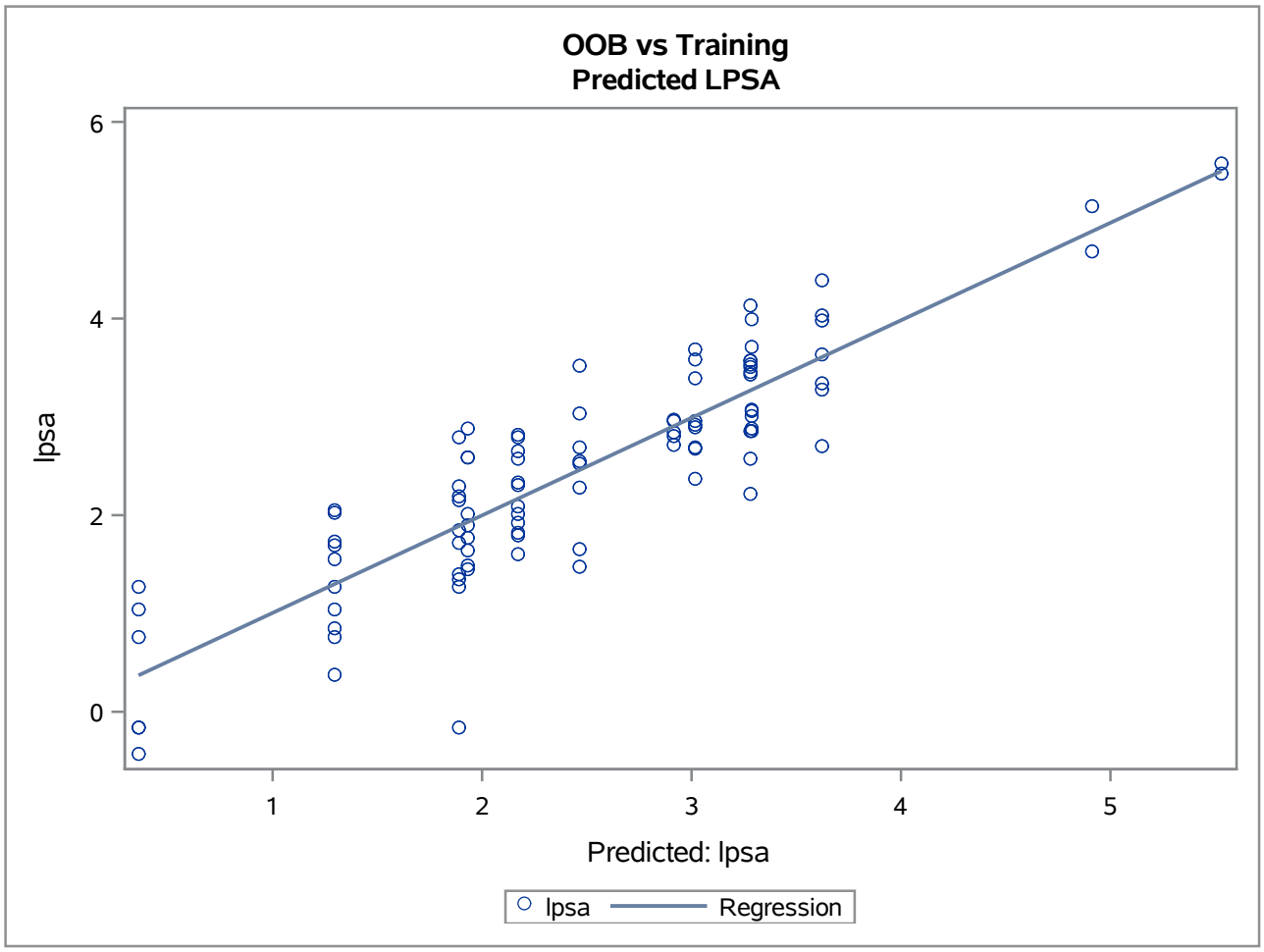
Fit Statistics			
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)
1	26	0.78620	1.10864
2	55	0.62269	1.08216
3	83	0.45109	0.95410
4	110	0.44811	0.96008
5	136	0.46758	0.97987
6	163	0.42888	0.92832
7	191	0.41425	0.86442
8	220	0.38271	0.80298
9	249	0.35752	0.75966
10	278	0.33867	0.69398
11	307	0.34620	0.69116
12	330	0.34941	0.68526
13	359	0.34208	0.68419
14	388	0.34275	0.68664
15	417	0.33352	0.66321
16	446	0.32522	0.65554
17	474	0.32617	0.65227
18	503	0.32373	0.64320
19	531	0.31054	0.62551
20	560	0.31257	0.62147
21	589	0.31712	0.61855
22	618	0.32107	0.62029
23	647	0.32337	0.61684
24	676	0.32585	0.61613
25	705	0.32325	0.61893
26	733	0.32635	0.61708
27	762	0.32707	0.61594
28	790	0.32020	0.61319
29	819	0.32015	0.60949

The HPFOREST Procedure

Fit Statistics			
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)
30	**848**	0.31307	0.60019
31	**877**	0.31293	0.60177
32	**906**	0.31174	0.59716
33	**935**	0.31031	0.59559
34	**963**	0.30808	0.59186
35	**991**	0.30884	0.59719
36	**1020**	0.30683	0.59495
37	**1049**	0.30671	0.59489
38	**1078**	0.30260	0.59141
39	**1107**	0.30313	0.59262
40	**1136**	0.30531	0.59698

Loss Reduction Variable Importance					
Variable	Number of Rules	MSE	OOB MSE	Absolute Error	OOB Absolute Error
lcavol	212	0.431118	0.20620	0.239354	0.067954
svi	41	0.089963	0.05895	0.037939	0.015437
lcp	120	0.205360	0.02845	0.105269	0.003971
gleason	25	0.080705	0.02155	0.042401	0.009424
pgg45	198	0.146063	-0.01492	0.121773	-0.003620
lweight	262	0.188958	-0.03879	0.180122	-0.016480
lbph	116	0.081953	-0.06141	0.080125	-0.027620
age	122	0.105200	-0.09742	0.084776	-0.031965


OOB vs Training
Average Squared Error
1.0
0.8
0.6
0.4
0
10
20
30
40
Number of Trees
Average Square Error (Full Data)
Average Square Error (OOB)


OOB vs Training
Predicted LPSA
lpsa
6
4
2
0
1
2
3
4
5
Predicted: lpsa
lpsa
Regression

The HPSPLIT Procedure

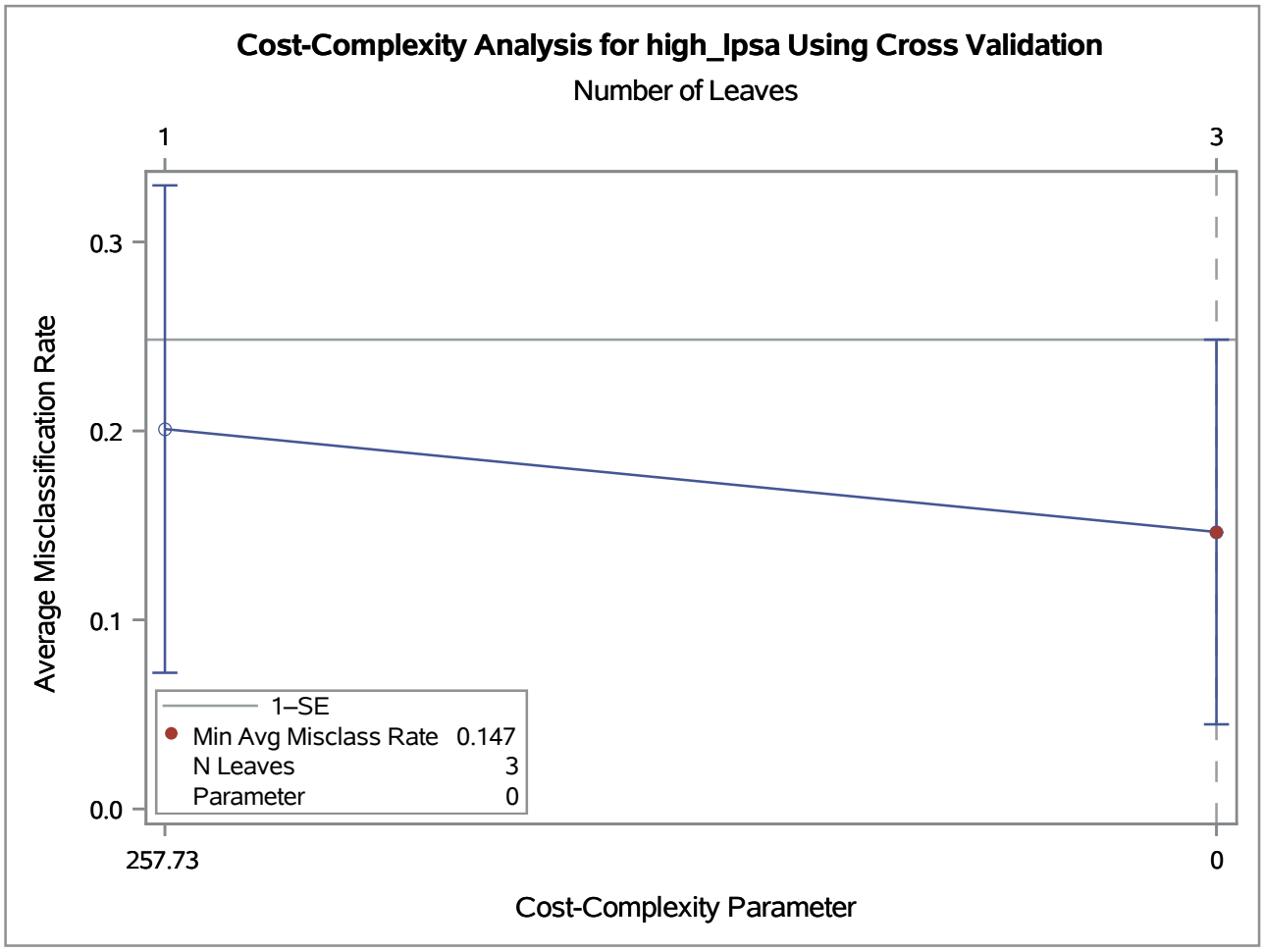
Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	**Engine**	**Role**	**Path**
WORK.PROSTRATE	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	3
Tree Depth	2
Number of Leaves Before Pruning	5
Number of Leaves After Pruning	3
Model Event Level	1

Number of Observations Read	97
Number of Observations Used	97

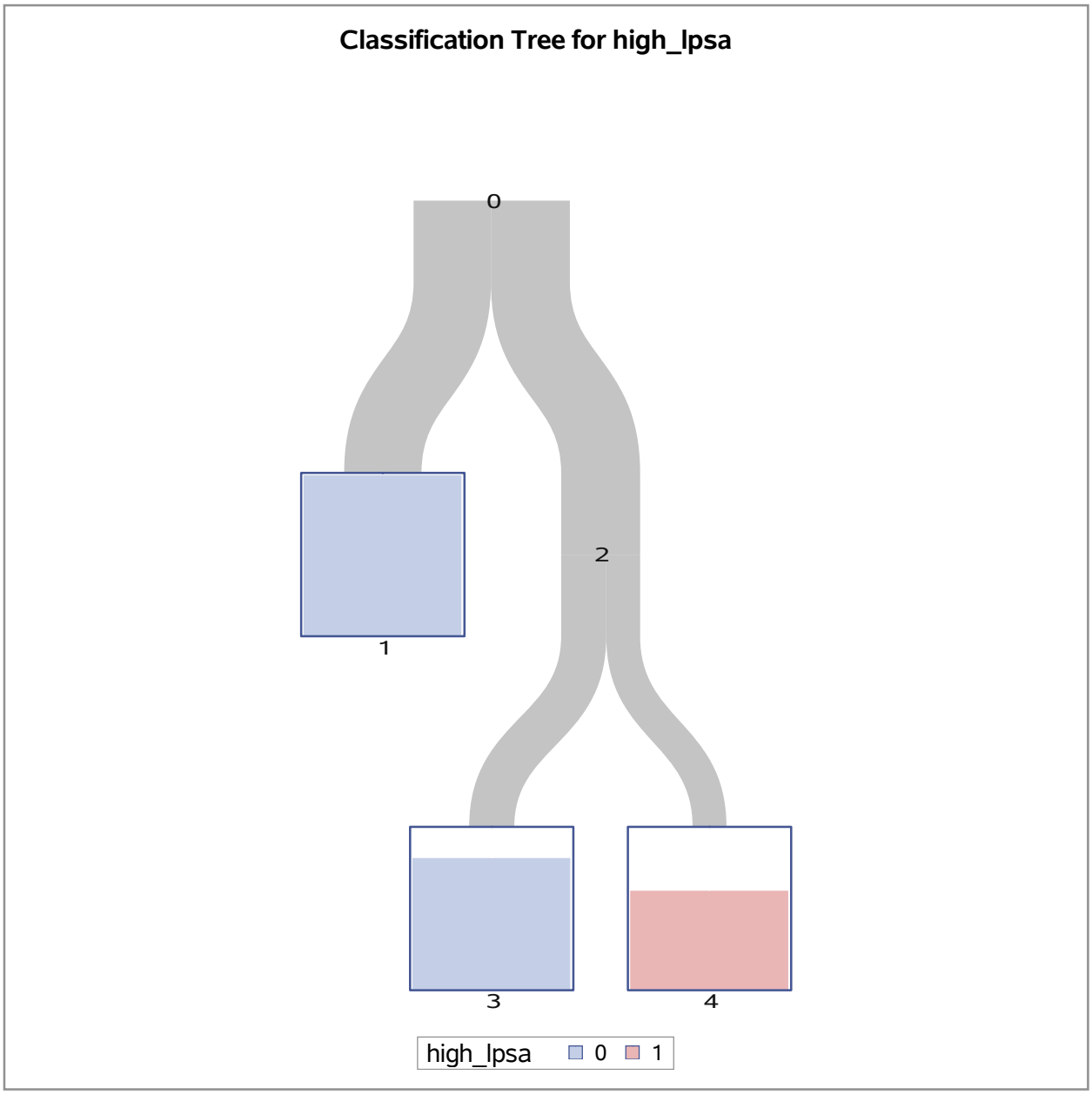
The HPSPLIT Procedure



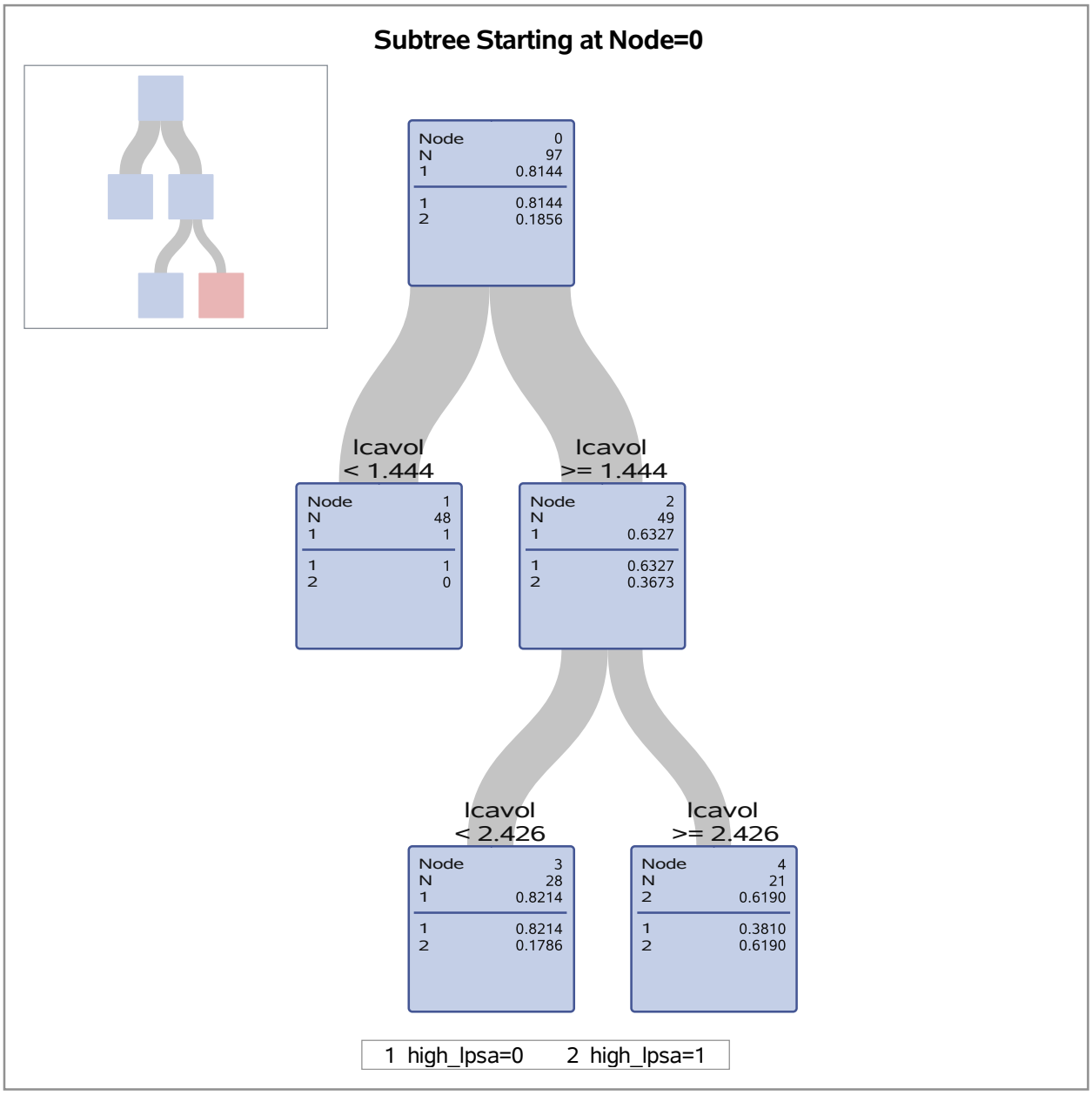
10-Fold Cross Validation Assessment of Model											
	Average Square Error				Number of Leaves			Misclassification Rate			
N Leaves	Min	Avg	Standard Error	Max	Min	Median	Max	Min	Avg	Standard Error	Max
3	0.00482	0.1215	0.0683	0.2211	2	3.0	3	0.0000	0.1742	0.1217	0.3333

10-Fold Cross Validation Confusion Matrix			
	Predicted		
Actual	0	1	Error Rate
0	70	9	0.1139
1	8	10	0.4444

The HPSPLIT Procedure



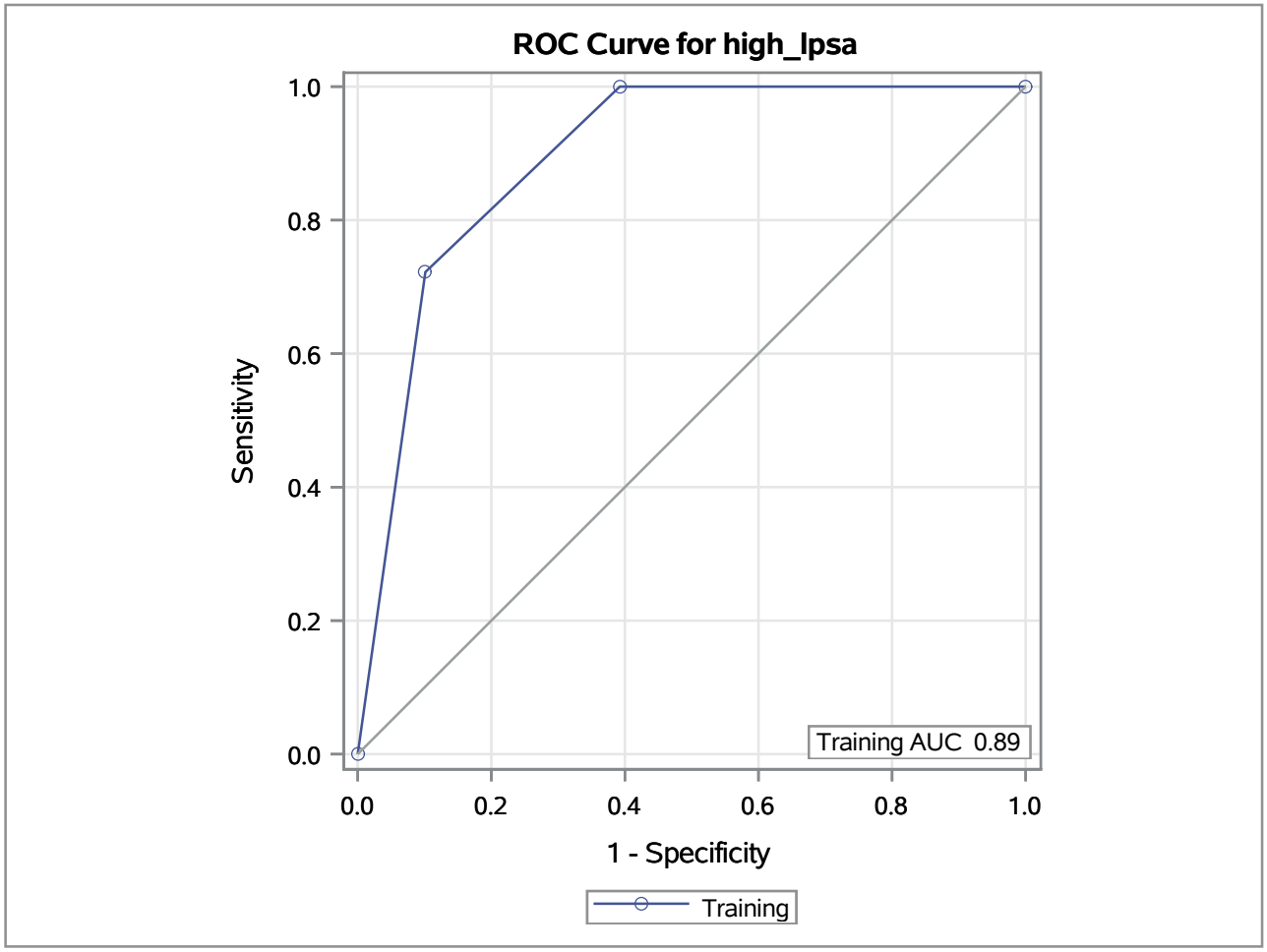
The HPSPLIT Procedure



The HPSPLIT Procedure

Confusion Matrices				
		Predicted		
	Actual	0	1	Error Rate
Model Based	0	71	8	0.1013
	1	5	13	0.2778
Cross Validation	0	70	9	0.1139
	1	8	10	0.4444

Fit Statistics for Selected Tree									
	N Leaves	ASE	Mis-class	Sensitivity	Specificity	Entropy	Gini	RSS	AUC
Model Based	3	0.0934	0.1340	0.7222	0.8987	0.4030	0.1868	18.1190	0.8949
Cross Validation	3	0.1215	0.1742	0.5556	0.8861				



The HPSPLIT Procedure

Variable Importance				
		Training		
Variable	Variable Label	Relative	Importance	Count
lcavol	lcavol	1.0000	3.3467	2

The HPFOREST Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	**Engine**	**Role**	**Path**
WORK.PROSTRATE	V9	Input	On Client
LDATA.SCORE	V9	Output	On Client

Model Information		
Parameter	**Value**	
Variables to Try	3	(Default)
Maximum Trees	30	
Actual Trees	30	
Inbag Fraction	0.5	
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	97
Number of Observations Used	97

The HPFOREST Procedure

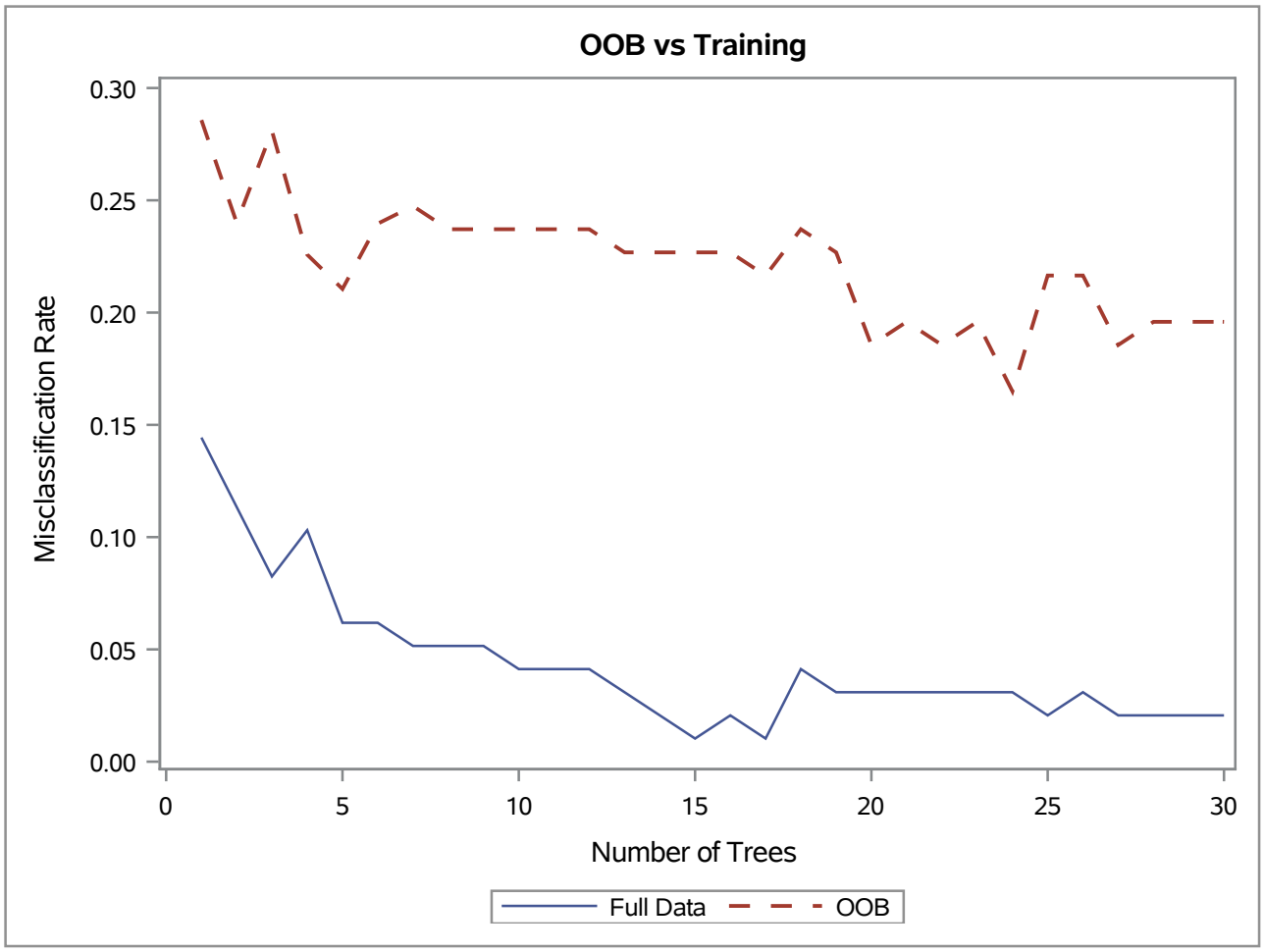
Baseline Fit Statistics	
Statistic	Value
Average Square Error	0.151
Misclassification Rate	0.186
Log Loss	0.480

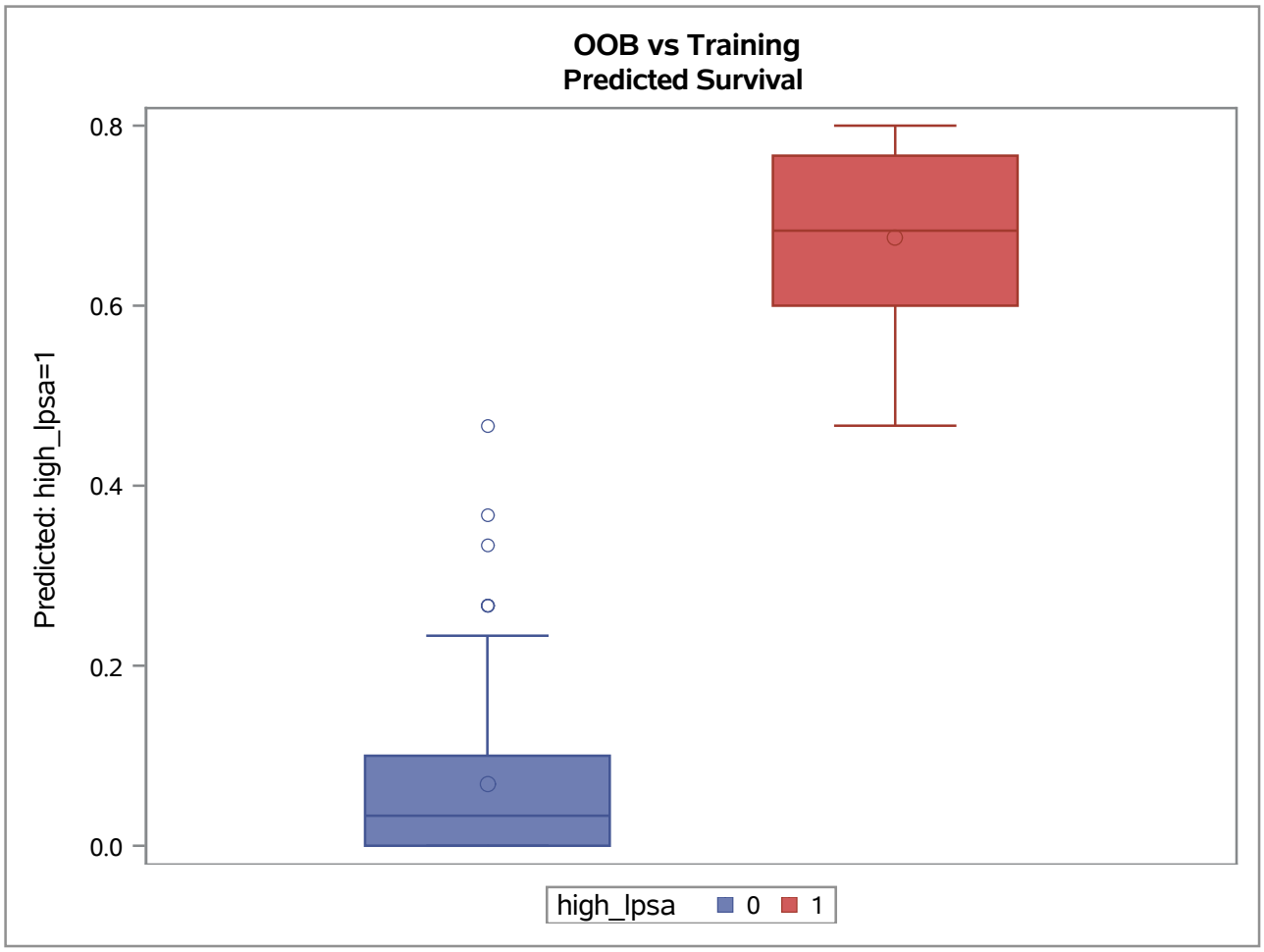
Fit Statistics							
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)	Misclassification Rate (Train)	Misclassification Rate (OOB)	Log Loss (Train)	Log Loss (OOB)
1	8	0.1443	0.286	0.1443	0.286	3.323	6.579
2	19	0.0799	0.243	0.1134	0.240	0.848	5.265
3	27	0.0619	0.243	0.0825	0.281	0.183	4.997
4	39	0.0593	0.201	0.1031	0.226	0.178	3.827
5	48	0.0524	0.189	0.0619	0.211	0.167	3.300
6	59	0.0470	0.165	0.0619	0.240	0.157	2.357
7	71	0.0457	0.177	0.0515	0.247	0.158	2.584
8	80	0.0441	0.176	0.0515	0.237	0.155	2.371
9	89	0.0410	0.170	0.0515	0.237	0.149	2.145
10	95	0.0411	0.166	0.0412	0.237	0.149	2.135
11	105	0.0408	0.163	0.0412	0.237	0.150	1.918
12	113	0.0420	0.162	0.0412	0.237	0.155	1.706
13	123	0.0412	0.165	0.0309	0.227	0.156	1.716
14	133	0.0403	0.159	0.0206	0.227	0.154	1.282
15	146	0.0368	0.146	0.0103	0.227	0.146	0.829
16	159	0.0356	0.145	0.0206	0.227	0.145	0.827
17	166	0.0355	0.144	0.0103	0.216	0.144	0.825
18	177	0.0353	0.144	0.0412	0.237	0.143	0.825
19	183	0.0360	0.138	0.0309	0.227	0.144	0.813
20	194	0.0351	0.136	0.0309	0.186	0.141	0.809
21	202	0.0346	0.133	0.0309	0.196	0.140	0.803
22	212	0.0348	0.130	0.0309	0.186	0.139	0.587
23	220	0.0356	0.131	0.0309	0.196	0.141	0.590
24	229	0.0346	0.128	0.0309	0.165	0.138	0.582
25	238	0.0340	0.127	0.0206	0.216	0.138	0.581
26	247	0.0343	0.126	0.0309	0.216	0.138	0.575
27	255	0.0343	0.126	0.0206	0.186	0.139	0.573
28	268	0.0335	0.125	0.0206	0.196	0.138	0.570

The HPFOREST Procedure

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)
29	**278**	0.0333	0.125	0.0206	0.196	0.138	0.569
30	**285**	0.0336	0.125	0.0206	0.196	0.138	0.569

Loss Reduction Variable Importance					
Variable	Number of Rules	Gini	OOB Gini	Margin	OOB Margin
svi	20	0.028868	0.00588	0.057735	0.03409
gleason	1	0.000988	-0.00045	0.001975	0.00060
lcp	40	0.055645	-0.00699	0.111290	0.04450
lcavol	50	0.075192	-0.00737	0.150383	0.07620
lbph	25	0.023908	-0.01951	0.047816	0.00509
age	25	0.022353	-0.02307	0.044707	-0.00195
pgg45	40	0.036801	-0.03365	0.073602	0.00303
lweight	54	0.044035	-0.04629	0.088071	-0.00520


OOB vs Training
Misclassification Rate
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25
30
Number of Trees
Full Data
OOB


OOB vs Training
Predicted Survival
Predicted: high_lpsa=1
0.0
0.2
0.4
0.6
0.8
high_lpsa
0
1

OOB vs Training
Predicted Survival

The FREQ Procedure

Frequency Col Pct

Table of high_lpsa by pred			
	pred		
high_lpsa	0	1	Total
0	79 97.53	0 0.00	79
1	2 2.47	16 100.00	18
Total	81	16	97

Statistics for Table of high_lpsa by pred

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	9.0000	2.4365	33.2438
One or more statistics not computed -- zero cell.			

Sample Size = 97

OOB vs Training Predicted Survival

The LOGISTIC Procedure

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

Number of Observations Read	97
Number of Observations Used	97

Response Profile		
Ordered Value	**high_lpsa**	**Total Frequency**
1	0	79
2	1	18

Probability modeled is high_lpsa=1.

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	95.068	64.658
SC	97.642	72.382
-2 Log L	93.068	58.658

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	34.4099	2	<.0001
Score	27.0141	2	<.0001
Wald	16.7160	2	0.0002

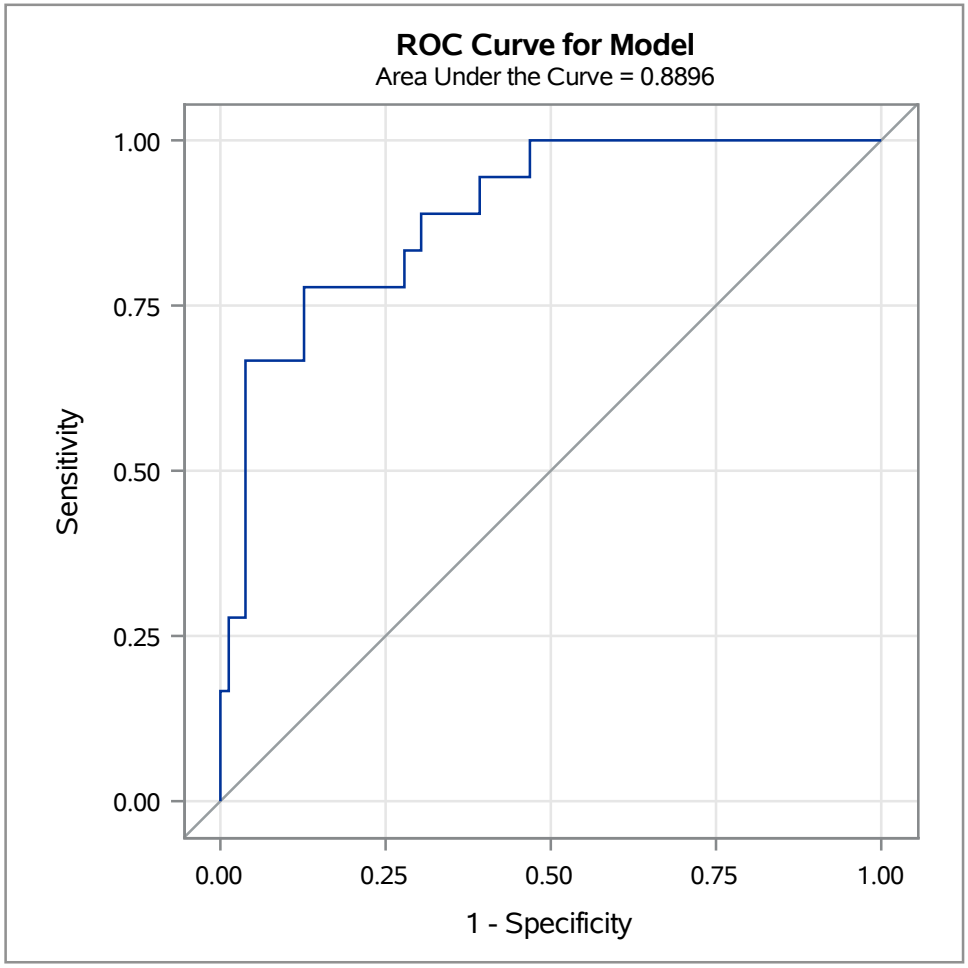
OOB vs Training
Predicted Survival

The LOGISTIC Procedure

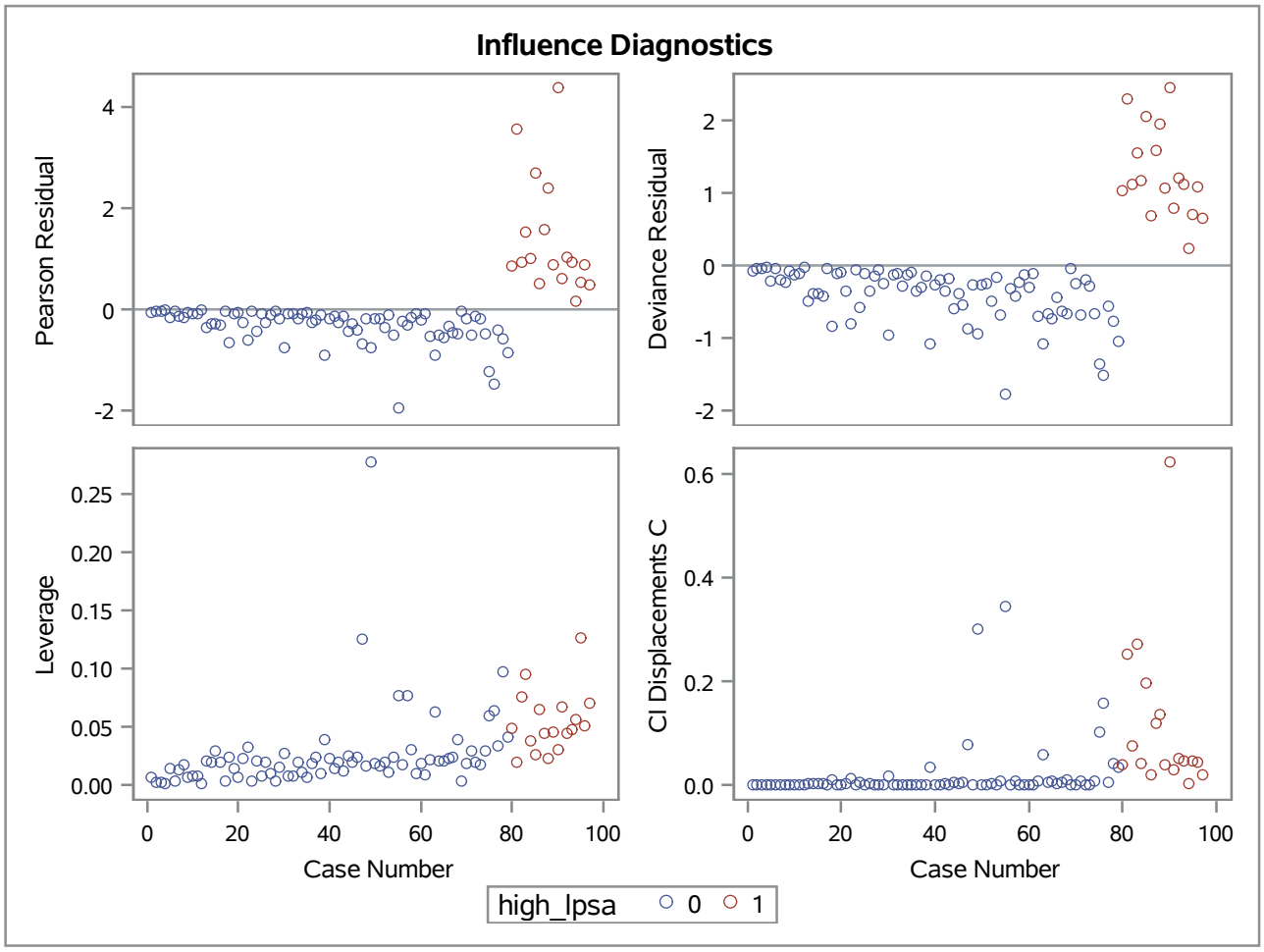
Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	-1.5398	3.1370	0.2410	0.6235	0.214
age	1	-0.0609	0.0498	1.4911	0.2220	0.941
lcavol	1	2.0543	0.5027	16.7017	<.0001	7.801

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
age	0.941	0.853	1.038
lcavol	7.801	2.913	20.894

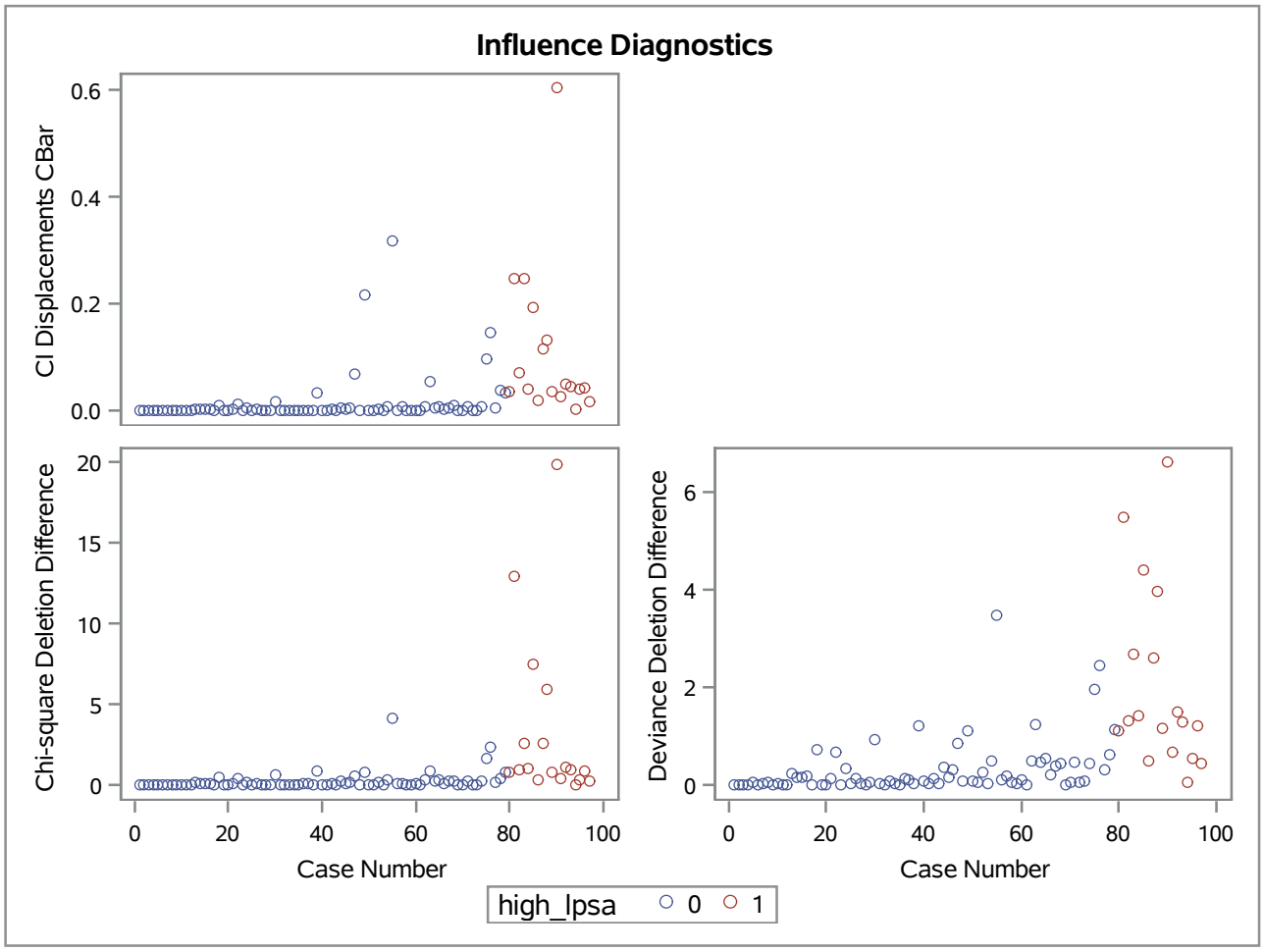
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.0	Somers' D	0.779
Percent Discordant	11.0	Gamma	0.779
Percent Tied	0.0	Tau-a	0.238
Pairs	1422	c	0.890



The LOGISTIC Procedure

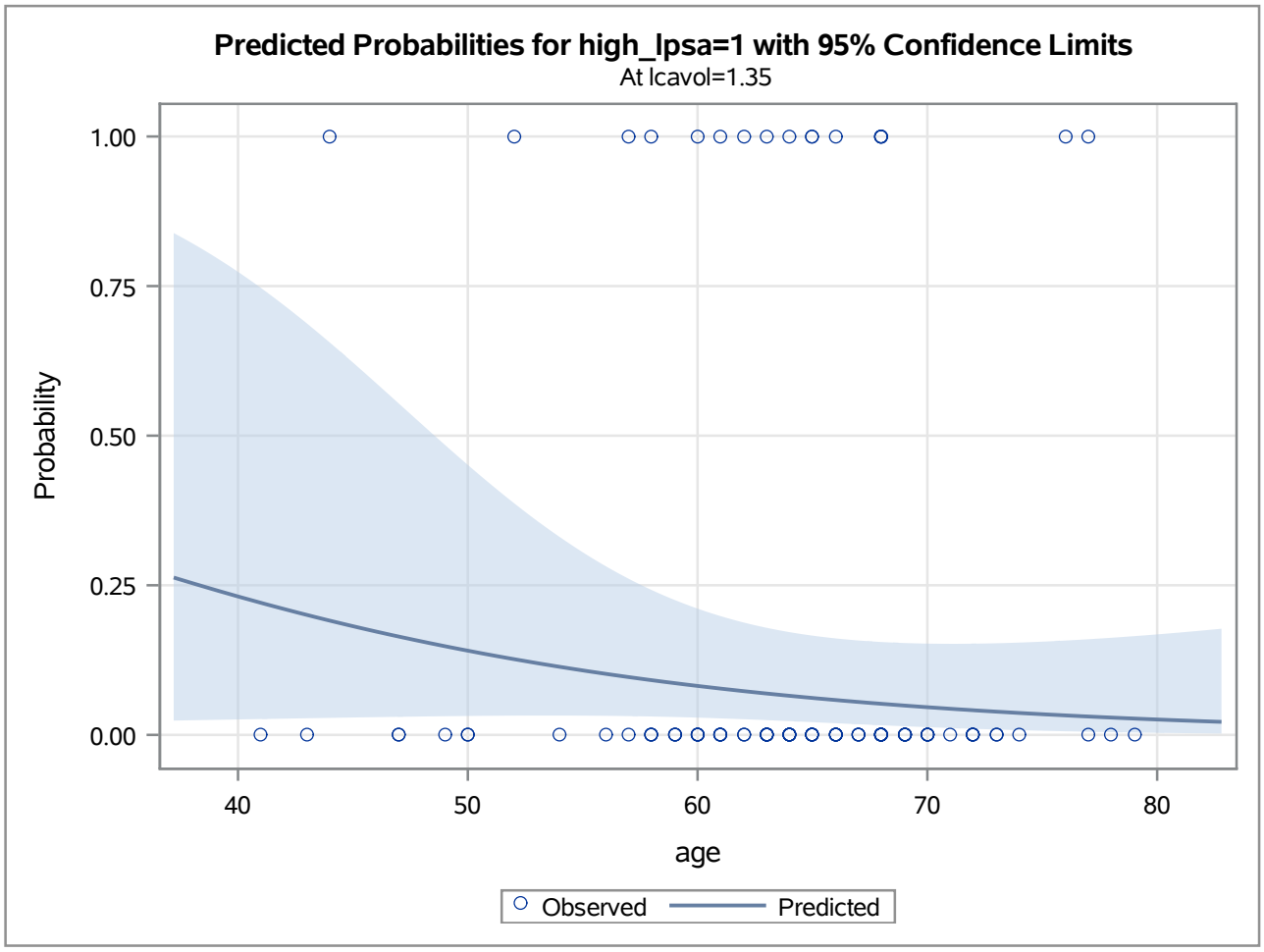


The LOGISTIC Procedure



OOB vs Training
Predicted Survival

The LOGISTIC Procedure



OOB vs Training
Predicted Survival

The FREQ Procedure

Frequency Col Pct	Table of high_lpsa by pred		
	pred(Estimated Probability)		
high_lpsa	0	1	Total
0	76 91.57	3 21.43	79
1	7 8.43	11 78.57	18
Total	83	14	97

Statistics for Table of high_lpsa by pred

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	39.8095	8.9451	177.1685
Relative Risk (Column 1)	2.4738	1.3840	4.4216
Relative Risk (Column 2)	0.0621	0.0193	0.2001

Sample Size = 97