

## Modifying SAS 9 Code for use in CAS through Viya

### PROC LOGISTIC EXAMPLE (SAS 9)

```
In [1]: title 'PROC LOGISTIC: Modeling with Categorical Predictors';

data Neuralgia;
  input Treatment $ Sex $ Age Duration Pain $ @@;
  datalines;
P F 68 1 No B M 74 16 No P F 67 30 No
P M 66 26 Yes B F 67 28 No B F 77 16 No
A F 71 12 No B F 72 50 No B F 76 9 Yes
A M 71 17 Yes A F 63 27 No A F 69 18 Yes
B F 66 12 No A M 62 42 No P F 64 1 Yes
A F 64 17 No P M 74 4 No A F 72 25 No
P M 70 1 Yes B M 66 19 No B M 59 29 No
A F 64 30 No A M 70 28 No A M 69 1 No
B F 78 1 No P M 83 1 Yes B F 69 42 No
B M 75 30 Yes P M 77 29 Yes P F 79 20 Yes
A M 70 12 No A F 69 12 No B F 65 14 No
B M 70 1 No B M 67 23 No A M 76 25 Yes
P M 78 12 Yes B M 77 1 Yes B F 69 24 No
P M 66 4 Yes P F 65 29 No P M 60 26 Yes
A M 78 15 Yes B M 75 21 Yes A F 67 11 No
P F 72 27 No P F 70 13 Yes A M 75 6 Yes
B F 65 7 No P F 68 27 Yes P M 68 11 Yes
P M 67 17 Yes B M 70 22 No A M 65 15 No
P F 67 1 Yes A M 67 10 No P F 72 11 Yes
A F 74 1 No B M 80 21 Yes A F 69 3 No
;

proc logistic data=Neuralgia;
  class Treatment Sex;
  model Pain= Treatment Sex Treatment*Sex Age Duration / expb;
run;

proc logistic data=Neuralgia;
  class Treatment Sex;
  model Pain=Treatment|Sex@2 Age Duration
    /selection=forward expb;
run;

ods graphics on;
proc logistic data=Neuralgia plots(only)=(oddsratio(range=clip));
  class Treatment Sex /param=ref;
  model Pain= Treatment Sex Age / noor;
  oddsratio Treatment;
  oddsratio Sex;
  oddsratio Age;
  contrast 'Pairwise A vs P' Treatment 1 0 / estimate=exp;
  contrast 'Pairwise B vs P' Treatment 0 1 / estimate=exp;
  contrast 'Pairwise A vs B' Treatment 1 -1 / estimate=exp;
  contrast 'Female vs Male' Sex 1 / estimate=exp;
  effectplot / at(Sex=all) noobs;
```

```
effectplot slicefit(sliceby=Sex plotby=Treatment) / noobs;  
run;
```

SAS server started using Context SAS Studio compute context with SESSION\_ID=549a896f-a1a8-4a32-833b-44b9c1aebd20-ses0000

## PROC LOGISTIC: Modeling with Categorical Predictors

### The LOGISTIC Procedure

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

Number of Observations Read	60
Number of Observations Used	60

Response Profile		
Ordered Value	Pain	Total Frequency
1	No	35
2	Yes	25

Probability modeled is Pain='No'.

Class Level Information			
Class	Value	Design Variables	
Treatment	A	1	0
	B	0	1
	P	-1	-1
Sex	F	1	
	M	-1	

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	83.503	64.596
SC	85.598	81.351
-2 Log L	81.503	48.596

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	32.9074	7	<.0001
Score	25.6812	7	0.0006
Wald	14.2879	7	0.0463

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Treatment	2	11.9886	0.0025
Sex	1	5.3104	0.0212
Treatment*Sex	2	0.1412	0.9318
Age	1	7.2744	0.0070
Duration	1	0.0247	0.8752

**Note:** Under full-rank parameterizations, Type 3 effect tests are replaced by joint tests. The joint test for an effect is a test that all the parameters associated with that effect are zero. Such joint tests might not be equivalent to Type 3 effect tests under GLM parameterization.

Analysis of Maximum Likelihood Estimates								
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept			1	19.2236	7.1315	7.2661	0.0070	2.232E8
Treatment	A		1	0.8483	0.5502	2.3773	0.1231	2.336
Treatment	B		1	1.4949	0.6622	5.0956	0.0240	4.459
Sex	F		1	0.9173	0.3981	5.3104	0.0212	2.503
Treatment*Sex	A	F	1	-0.2010	0.5568	0.1304	0.7180	0.818
Treatment*Sex	B	F	1	0.0487	0.5563	0.0077	0.9302	1.050
Age			1	-0.2688	0.0996	7.2744	0.0070	0.764
Duration			1	0.00523	0.0333	0.0247	0.8752	1.005

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Age	0.764	0.629	0.929
Duration	1.005	0.942	1.073

Association of Predicted Probabilities and Observed Responses

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.5	Somers' D	0.810
Percent Discordant	9.5	Gamma	0.810
Percent Tied	0.0	Tau-a	0.401
Pairs	875	c	0.905

## PROC LOGISTIC: Modeling with Categorical Predictors

### The LOGISTIC Procedure

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

Number of Observations Read	60
Number of Observations Used	60

Response Profile		
Ordered Value	Pain	Total Frequency
1	No	35
2	Yes	25

Probability modeled is Pain='No'.

### Forward Selection Procedure

Class Level Information			
Class	Value	Design Variables	
Treatment	A	1	0
	B	0	1
	P	-1	-1
Sex	F	1	
	M	-1	

Step 0. Intercept entered:

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

<b>-2 Log L</b>	=	81.503
-----------------	---	--------

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
25.6812	7	0.0006

## Step 1. Effect Treatment entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
<b>AIC</b>	83.503	73.480
<b>SC</b>	85.598	79.763
<b>-2 Log L</b>	81.503	67.480

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
<b>Likelihood Ratio</b>	14.0230	2	0.0009
<b>Score</b>	13.7143	2	0.0011
<b>Wald</b>	12.0695	2	0.0024

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
15.5126	5	0.0084

## Step 2. Effect Age entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
<b>AIC</b>	83.503	63.044

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
SC	85.598	71.421
-2 Log L	81.503	55.044

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.4591	3	<.0001
Score	21.8943	3	<.0001
Wald	14.1262	3	0.0027

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
6.1267	4	0.1899

### Step 3. Effect Sex entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	83.503	58.767
SC	85.598	69.239
-2 Log L	81.503	48.767

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	32.7358	4	<.0001
Score	25.6611	4	<.0001
Wald	14.5666	4	0.0057

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.1734	3	0.9818

**Note:** No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Forward Selection					
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	Treatment	2	1	13.7143	0.0011
2	Age	1	2	10.6038	0.0011
3	Sex	1	3	5.9959	0.0143

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Treatment	2	12.6928	0.0018
Sex	1	5.3013	0.0213
Age	1	7.6314	0.0057

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	19.0804	6.7882	7.9007	0.0049	1.9343E8
Treatment	A	1	0.8772	0.5274	2.7662	0.0963	2.404
Treatment	B	1	1.4246	0.6036	5.5711	0.0183	4.156
Sex	F	1	0.9118	0.3960	5.3013	0.0213	2.489
Age		1	-0.2650	0.0959	7.6314	0.0057	0.767

Odds Ratio Estimates				
Effect	Point Estimate	95% Wald Confidence Limits		
Treatment A vs P	24.022	3.295	175.121	
Treatment B vs P	41.528	4.500	383.262	
Sex F vs M	6.194	1.312	29.248	
Age	0.767	0.636	0.926	

Association of Predicted Probabilities and Observed Responses				
Percent Concordant	90.3	Somers' D	0.811	
Percent Discordant	9.1	Gamma	0.816	
Percent Tied	0.6	Tau-a	0.401	
Pairs	875	c	0.906	



## The LOGISTIC Procedure

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

Number of Observations Read	60
Number of Observations Used	60

Response Profile		
Ordered Value	Pain	Total Frequency
1	No	35
2	Yes	25

Probability modeled is Pain='No'.

Class Level Information			
Class	Value	Design Variables	
Treatment	A	1	0
	B	0	1
	P	0	0
Sex	F	1	
	M	0	

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	83.503	58.767
SC	85.598	69.239
-2 Log L	81.503	48.767

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq

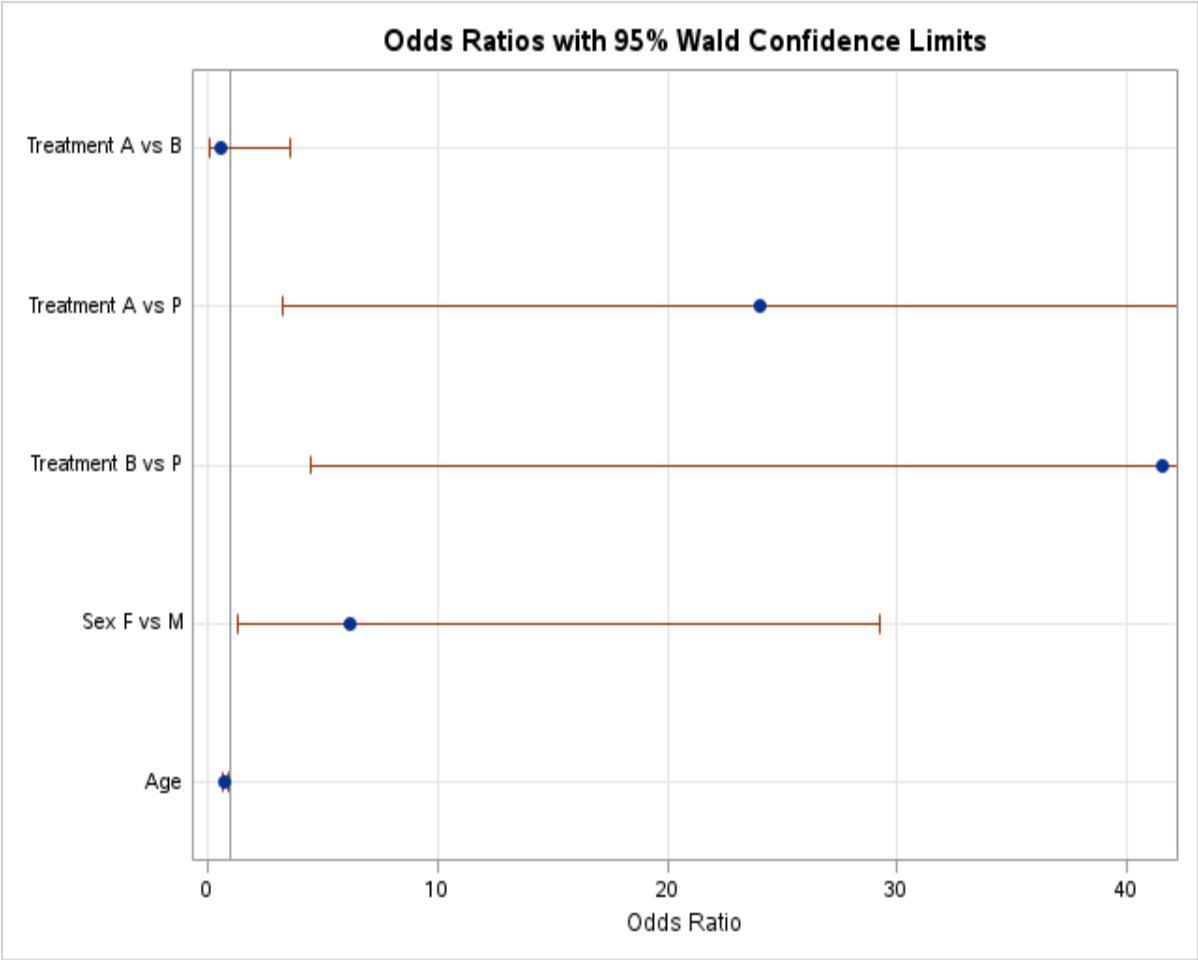
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	32.7358	4	<.0001
Score	25.6611	4	<.0001
Wald	14.5666	4	0.0057

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Treatment	2	12.6928	0.0018
Sex	1	5.3013	0.0213
Age	1	7.6314	0.0057

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	15.8669	6.4056	6.1357	0.0132
Treatment	A	1	3.1790	1.0135	9.8375	0.0017
Treatment	B	1	3.7264	1.1339	10.8006	0.0010
Sex	F	1	1.8235	0.7920	5.3013	0.0213
Age		1	-0.2650	0.0959	7.6314	0.0057

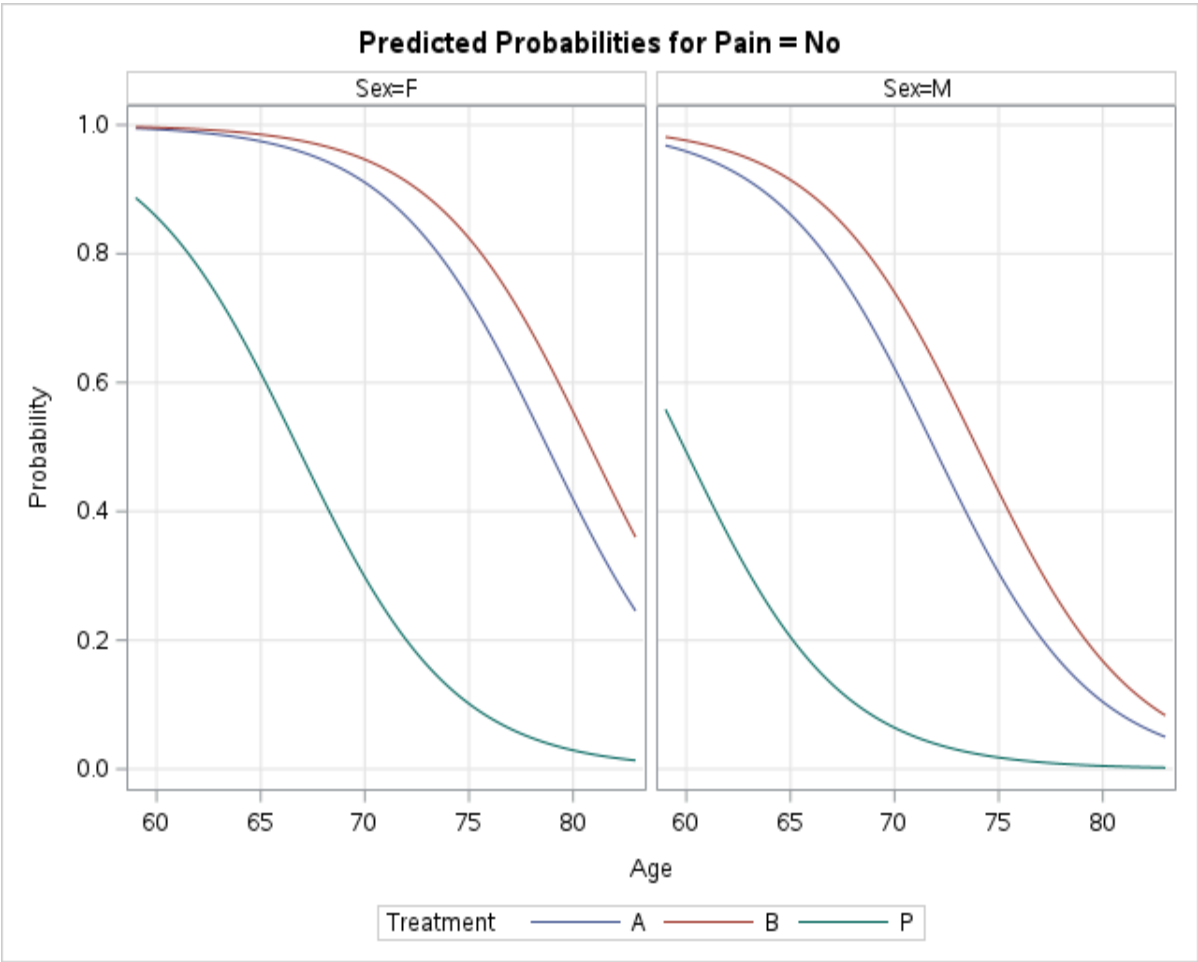
Association of Predicted Probabilities and Observed Responses				
Percent Concordant		90.3	Somers' D	0.811
Percent Discordant		9.1	Gamma	0.816
Percent Tied		0.6	Tau-a	0.401
Pairs		875	c	0.906

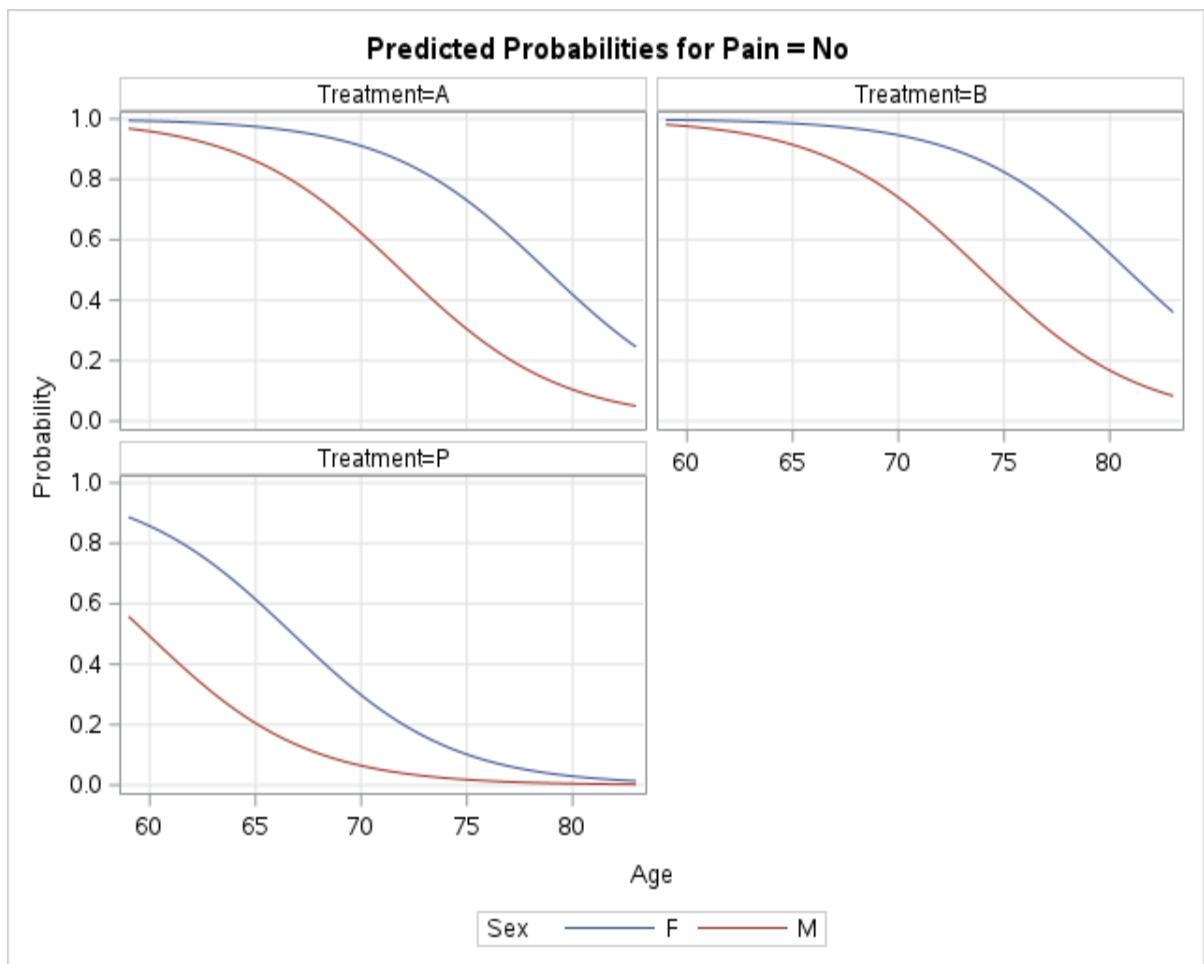
Odds Ratio Estimates and Wald Confidence Intervals			
Odds Ratio	Estimate	95% Confidence Limits	
Treatment A vs B	0.578	0.093	3.589
Treatment A vs P	24.022	3.295	175.121
Treatment B vs P	41.528	4.500	383.262
Sex F vs M	6.194	1.312	29.248
Age	0.767	0.636	0.926



Contrast Test Results			
Contrast	DF	Wald Chi-Square	Pr > ChiSq
Pairwise A vs P	1	9.8375	0.0017
Pairwise B vs P	1	10.8006	0.0010
Pairwise A vs B	1	0.3455	0.5567
Female vs Male	1	5.3013	0.0213

Contrast Estimation and Testing Results by Row									
Contrast	Type	Row	Estimate	Standard Error	Alpha	Confidence Limits		Wald Chi-Square	Pr > ChiSq
Pairwise A vs P	EXP	1	24.0218	24.3473	0.05	3.2951	175.1	9.8375	0.0017
Pairwise B vs P	EXP	1	41.5284	47.0877	0.05	4.4998	383.3	10.8006	0.0010
Pairwise A vs B	EXP	1	0.5784	0.5387	0.05	0.0932	3.5889	0.3455	0.5567
Female vs Male	EXP	1	6.1937	4.9053	0.05	1.3116	29.2476	5.3013	0.0213





PROC LOGSELECT (CAS enabled Procedure)

```
In [2]: cas mySession sessopts=(caslib=casuser timeout=1800 locale="en_US");
libname casuser cas;
```

```

68 ods listing close;ods html5 (id=saspy_internal) options(bitmap_mode='inline') de
vice=svg style=HTMLBlue; ods graphics on /
68 ! outputfmt=png;
NOTE: Writing HTML5(SASPY_INTERNAL) Body file: sashtml1.htm
69
70 cas mySession sessopts=(caslib=casuser timeout=1800 locale="en_US");
NOTE: The session MYSESSION connected successfully to Cloud Analytic Services sas-cas
-server-default-client using port 5570. The
      UUID is b2da7e84-a634-014b-a369-c5f8d1e66ee9. The user is danny.modlin@sas.com
and the active caslib is
      CASUSER(danny.modlin@sas.com).
NOTE: The SAS option SESSREF was updated with the value MYSESSION.
NOTE: The SAS macro _SESSREF_ was updated with the value MYSESSION.
NOTE: The session is using 0 workers.
NOTE: 'CASUSER(danny.modlin@sas.com)' is now the active caslib.
NOTE: The CAS statement request to update one or more session options for session MYSE
SSION completed.
71 libname casuser cas;
NOTE: Libref CASUSER was successfully assigned as follows:
      Engine:          CAS
      Physical Name: b2da7e84-a634-014b-a369-c5f8d1e66ee9
72 ods html5 (id=saspy_internal) close;ods listing;
73

```

```

In [3]: title 'PROC LOGSELECT: Modeling Binomial Data';

data Ingots;
  input Heat Soak r n @@;
  Obsnum= _n_;
  datalines;
7 1.0 0 10 14 1.0 0 31 27 1.0 1 56 51 1.0 3 13
7 1.7 0 17 14 1.7 0 43 27 1.7 4 44 51 1.7 0 1
7 2.2 0 7 14 2.2 2 33 27 2.2 0 21 51 2.2 0 1
7 2.8 0 12 14 2.8 0 31 27 2.8 1 22 51 4.0 0 1
7 4.0 0 9 14 4.0 0 19 27 4.0 1 16
;
data casuser.Ingots;
  set Ingots;
run;

proc logselect data=casuser.Ingots association ctable(out=casuser.Roc nocounts tpf fpf
model r/n = Heat Soak Heat*Soak;
  output out=casuser.Out xbeta predicted=Pred copyvars=(Heat Soak);
run;

proc print data=casuser.Out;
  where Heat=14 & Soak=1.7;
run;

ods graphics on;
proc sgplot data=casuser.Roc aspect=1 noautolegend;
  title 'ROC Curve';
  xaxis values=(0 to 1 by 0.25) grid offsetmin=.05 offsetmax=.05;
  yaxis values=(0 to 1 by 0.25) grid offsetmin=.05 offsetmax=.05;
  lineparm x=0 y=0 slope=1 / lineattrs=(color=ligr);

```

```

series x=FPF y=TPF;
inset 'Area under the curve=0.7706' / position=bottomright;
run;

data casuser.Ingots2;
set Ingots;
a = n - r;
run;

proc logselect data=casuser.Ingots2 association ctable(out=casuser.Roc nocounts tpf fp
model r/a = Heat Soak Heat*Soak;
output out=casuser.Out xbeta predicted=Pred copyvars=(Heat Soak);
run;

data casuser.Ingots_binary;
set Ingots;
do i=1 to n;
if i <= r then y=1; else y = 0;
output;
end;
run;

proc logselect data=casuser.Ingots_binary;
model y(event='1') = Heat Soak Heat*Soak;
run;

```

## PROC LOGSELECT: Modeling Binomial Data

### The LOGSELECT Procedure

Model Information	
Data Source	INGOTS
Response Variable (Events)	r
Response Variable (Trials)	n
Distribution	Binomial
Link Function	Logit
Optimization Technique	Newton-Raphson with Ridging

Number of Observations Read	19
Number of Observations Used	19

Response Profile		
Ordered Value	Binary Outcome	Total Frequency
1	Event	12
2	Nonevent	375

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

Dimensions	
Columns in Design	4
Number of Effects	4
Max Effect Columns	1
Rank of Design	4
Parameters in Optimization	4

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	3	11.7663	0.0082

Fit Statistics	
-2 Log Likelihood	27.95689
AIC (smaller is better)	35.95689
AICC (smaller is better)	38.81403



Fit Statistics	
<b>SBC (smaller is better)</b>	39.73464

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>Intercept</b>	1	-5.990191	1.666622	12.9183	0.0003
<b>Heat</b>	1	0.096339	0.047067	4.1896	0.0407
<b>Soak</b>	1	0.299574	0.755068	0.1574	0.6916
<b>Heat * Soak</b>	1	-0.008840	0.025319	0.1219	0.7270

Association of Predicted Probabilities and Observed Responses	
<b>Concordance Index (AUC)</b>	0.7706
<b>Somers' D</b>	0.5411
<b>Gamma</b>	0.5858
<b>Tau-a</b>	0.0326
<b>Pairs</b>	4500
<b>Percent Concordant</b>	73.2444
<b>Percent Discordant</b>	19.1333
<b>Percent Tied</b>	7.6222

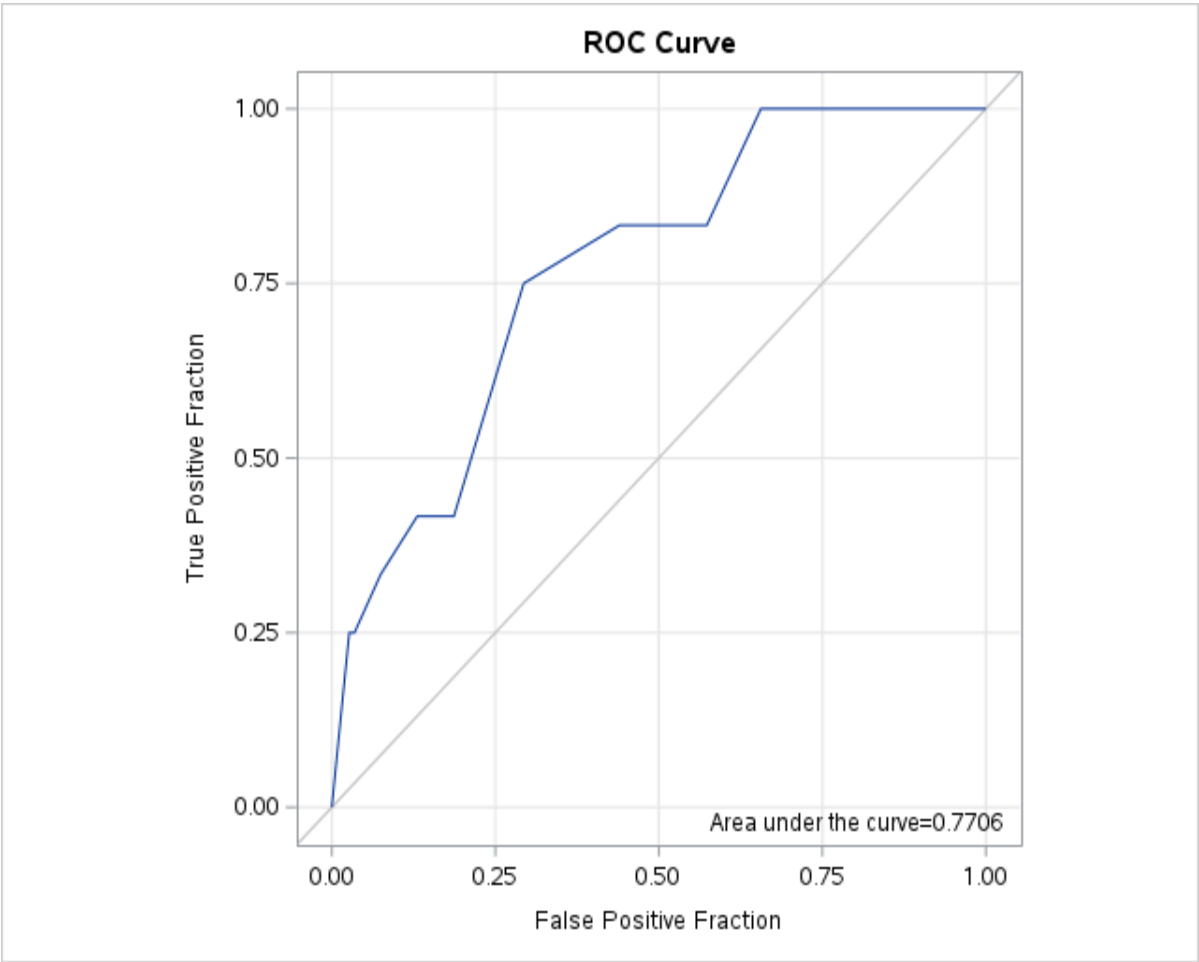
Task Timing		
Task	Seconds	Percent
<b>Setup and Parsing</b>	0.00	18.98%
<b>Levelization</b>	0.00	5.74%
<b>Model Initialization</b>	0.00	4.19%
<b>SSCP Computation</b>	0.00	4.79%
<b>Model Fitting</b>	0.01	50.85%
<b>Creating Output Data</b>	0.00	3.70%
<b>Association and Classification</b>	0.00	9.30%
<b>Display</b>	0.00	2.32%
<b>Cleanup</b>	0.00	2.26%
<b>Total</b>	0.02	100.00%

Output CAS Tables			
CAS Library	Name	Number of Rows	Number of Columns
<b>CASUSER(danny.modlin@sas.com)</b>	<b>OUT</b>	19	4

Output CAS Tables			
CAS Library	Name	Number of Rows	Number of Columns
CASUSER(danny.modlin@sas.com)	ROC	20	3

PROC LOGSELECT: Modeling Binomial Data

Obs	Pred	_XBETA_	Heat	Soak
1	0.012836	-4.34256	14	1.7



ROC Curve

The LOGSELECT Procedure

Model Information	
Data Source	INGOTS2

Model Information	
Response Variable (Events)	r
Response Variable (Trials)	a
Distribution	Binomial
Link Function	Logit
Optimization Technique	Newton-Raphson with Ridging

Number of Observations Read	19
Number of Observations Used	19

Response Profile		
Ordered Value	Binary Outcome	Total Frequency
1	Event	12
2	Nonevent	363

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

Dimensions	
Columns in Design	4
Number of Effects	4
Max Effect Columns	1
Rank of Design	4
Parameters in Optimization	4

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	3	13.3554	0.0039

Fit Statistics	
-2 Log Likelihood	28.64026
AIC (smaller is better)	36.64026
AICC (smaller is better)	39.49740
SBC (smaller is better)	40.41802

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-6.278612	1.678605	13.9904	0.0002
Heat	1	0.110619	0.048161	5.2757	0.0216
Soak	1	0.398510	0.750555	0.2819	0.5955
Heat * Soak	1	-0.013243	0.025440	0.2710	0.6027

Association of Predicted Probabilities and Observed Responses	
Concordance Index (AUC)	0.7795
Somers' D	0.5590
Gamma	0.6020
Tau-a	0.0347
Pairs	4356
Percent Concordant	74.3802
Percent Discordant	18.4803
Percent Tied	7.1396

Task Timing		
Task	Seconds	Percent
Setup and Parsing	0.00	18.70%
Levelization	0.00	6.22%
Model Initialization	0.00	4.13%
SSCP Computation	0.00	3.30%
Model Fitting	0.01	56.77%
Creating Output Data	0.00	4.21%
Association and Classification	0.00	4.40%
Display	0.00	2.15%
Cleanup	0.00	2.09%
Total	0.02	100.00%

Output CAS Tables			
CAS Library	Name	Number of Rows	Number of Columns
CASUSER(danny.modlin@sas.com)	OUT	19	4
CASUSER(danny.modlin@sas.com)	ROC	20	3

## ROC Curve

### The LOGSELECT Procedure

Model Information	
Data Source	INGOTS_BINARY
Response Variable	y
Distribution	Binary
Link Function	Logit
Optimization Technique	Newton-Raphson with Ridging

Number of Observations Read	387
Number of Observations Used	387

Response Profile		
Ordered Value	y	Total Frequency
1	0	375
2	1	12

Probability modeled is y = 1.

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

Dimensions	
Columns in Design	4
Number of Effects	4
Max Effect Columns	1
Rank of Design	4
Parameters in Optimization	4

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	3	11.7663	0.0082

Fit Statistics	
-2 Log Likelihood	95.22218
AIC (smaller is better)	103.22218
AICC (smaller is better)	103.32690

Fit Statistics	
<b>SBC (smaller is better)</b>	119.05588

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>Intercept</b>	1	-5.990191	1.666622	12.9183	0.0003
<b>Heat</b>	1	0.096339	0.047067	4.1896	0.0407
<b>Soak</b>	1	0.299574	0.755068	0.1574	0.6916
<b>Heat * Soak</b>	1	-0.008840	0.025319	0.1219	0.7270

Task Timing		
Task	Seconds	Percent
<b>Setup and Parsing</b>	0.00	20.15%
<b>Levelization</b>	0.00	7.27%
<b>Model Initialization</b>	0.00	4.09%
<b>SSCP Computation</b>	0.00	4.24%
<b>Model Fitting</b>	0.01	64.18%
<b>Display</b>	0.00	0.06%
<b>Cleanup</b>	0.00	0.01%
<b>Total</b>	0.02	100.00%

## Logistic Regression CAS Action

```
In [4]: proc cas;
dataStep.runCode /
  code="
    data getStarted;
      nTotalObs=1000;
      drop c2 eta pr i rew nTotalObs nObsPerThread nExtras;
      call streaminit(1);
      nObsPerThread = int(nTotalObs/_nthreads_);
      nExtras       = mod(nTotalObs,_nthreads_);
      if _threadid_ <= nExtras then nObsPerThread = nObsPerThread + 1;
      do i=1 to nObsPerThread;
        id = (_threadid_ - 1) * nObsPerThread + i;
        if _threadid_ > nExtras then id = id + nExtras;
        rew = rand('rewind', id);
        x1=round(rand('normal')*5+10,.1); x2=round(7*rand('uniform'))/7;
        x3=round(rand('normal')*1+2,.1);  x4=round(50*rand('uniform'));
        x5=round(100*rand('uniform'));     x6=round(rand('normal')*.8+1.5,.1);
        x7=10*round(10*rand('uniform'));   x8=round(10*rand('uniform'))/10;
        x9=round(rand('normal')*3+5,.1);   x10=round(rand('normal')*2+3,.1);
        c2=rand('uniform');
      end;
    end;
  end;
run;
```

```

        if (c2<.1) then C='A'; else if (c2<.2) then C='B';
        else if (c2<.3) then C='C'; else if (c2<.4) then C='D';
        else if (c2<.5) then C='E'; else if (c2<.6) then C='F';
        else if (c2<.7) then C='G'; else if (c2<.8) then C='H';
        else if (c2<.9) then C='I'; else C='J';
        eta=1-x2-x8;
        pr= exp(eta)/(1+exp(eta));
        y=(rand('uniform') > pr);
        output;
    end;
run;
",
single="no";
run;

proc cas;
regression.logistic /
    class={"C"},
    model={depvar="y",
        effects={"C", "x1", "x2", "x3", "x4", "x5", "x6", "x7", "x8", "x9",
            "x10"}},
    optimization={itHist="summary"},
    outputTables={names={parameterestimates="pe"}},
    table="getStarted";
run;

proc cas;
regression.logistic /
    class={"C"},
    display={traceNames="true"},
    model={depvar="y",
        effects={"C", "x1", "x2", "x3", "x4", "x5", "x6", "x7", "x8", "x9",
            "x10"}},
    selection={details="all", method="forward"},
    table="getStarted";
run;

```

## ROC Curve

### Results from dataStep.runCode

Output CAS Tables			
CAS Library	Name	Number of Rows	Number of Columns
CASUSER(danny.modlin@sas.com)	getStarted	1000	13

## ROC Curve

### Results from regression.logistic

Model Information	
Data Source	GETSTARTED
Response Variable	y
Distribution	Binary
Link Function	Logit
Optimization Technique	Newton-Raphson with Ridging

Number of Observations Read	1000
Number of Observations Used	1000

Response Profile		
Ordered Value	y	Total Frequency
1	0	484
2	1	516

Probability modeled is y = 0.

Class Level Information		
Class	Levels	Values
C	10	A B C D E F G H I J

Iteration History				
Iteration	Evaluations	Objective Function	Change	Maximum Gradient
0	4	0.6613855981		0.20692
1	2	0.6583727872	0.00301281	0.01883



Iteration History				
Iteration	Evaluations	Objective Function	Change	Maximum Gradient
2	2	0.6583596409	0.00001315	0.000082
3	2	0.6583596407	0.00000000	1.23E-9

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

Dimensions	
Columns in Design	21
Number of Effects	12
Max Effect Columns	10
Rank of Design	20
Parameters in Optimization	20

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	19	68.5509	<.0001

Fit Statistics	
-2 Log Likelihood	1316.71928
AIC (smaller is better)	1356.71928
AICC (smaller is better)	1357.57730
SBC (smaller is better)	1454.87439

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.091372	0.419472	0.0474	0.8276
C A	1	0.101855	0.295610	0.1187	0.7304
C B	1	0.313845	0.289155	1.1781	0.2778
C C	1	0.514901	0.288989	3.1746	0.0748
C D	1	0.190620	0.307220	0.3850	0.5350
C E	1	0.115930	0.285505	0.1649	0.6847
C F	1	0.488200	0.292348	2.7887	0.0949
C G	1	0.607139	0.290986	4.3534	0.0369
C H	1	0.422393	0.286422	2.1748	0.1403
C I	1	0.099037	0.284288	0.1214	0.7276

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>C J</b>	0	0	.	.	.
<b>x1</b>	1	0.000629	0.013073	0.0023	0.9616
<b>x2</b>	1	-1.133344	0.228116	24.6838	<.0001
<b>x3</b>	1	0.077254	0.065331	1.3983	0.2370
<b>x4</b>	1	0.001466	0.004652	0.0993	0.7526
<b>x5</b>	1	0.003207	0.002301	1.9414	0.1635
<b>x6</b>	1	0.041222	0.083063	0.2463	0.6197
<b>x7</b>	1	-0.001533	0.002237	0.4694	0.4933
<b>x8</b>	1	-1.063694	0.232968	20.8469	<.0001
<b>x9</b>	1	0.015834	0.022353	0.5018	0.4787
<b>x10</b>	1	0.074454	0.033162	5.0408	0.0248

Task Timing		
Task	Seconds	Percent
<b>Setup and Parsing</b>	0.01	14.21%
<b>Levelization</b>	0.00	4.06%
<b>Model Initialization</b>	0.00	2.64%
<b>SSCP Computation</b>	0.02	48.68%
<b>Model Fitting</b>	0.01	30.37%
<b>Display</b>	0.00	0.03%
<b>Cleanup</b>	0.00	0.00%
<b>Total</b>	0.04	100.00%

Output CAS Tables			
CAS Library	Name	Number of Rows	Number of Columns
CASUSER(danny.modlin@sas.com)	pe	21	9

## ROC Curve

## Results from regression.logistic

Model Information	
Data Source	GETSTARTED

Model Information	
Response Variable	y
Distribution	Binary
Link Function	Logit
Optimization Technique	Newton-Raphson with Ridging

Number of Observations Read	1000
Number of Observations Used	1000

Response Profile		
Ordered Value	y	Total Frequency
1	0	484
2	1	516

Probability modeled is y = 0.

Class Level Information		
Class	Levels	Values
C	10	A B C D E F G H I J

Selection Information	
Selection Method	Forward
Select Criterion	SBC
Stop Criterion	SBC
Effect Hierarchy Enforced	None
Stop Horizon	3

Forward Selection: Step 0

Effects Included: Intercept

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

Dimensions	
Columns in Design	1
Number of Effects	1
Max Effect Columns	1
Rank of Design	1

Dimensions	
Parameters in Optimization	1

Fit Statistics	
-2 Log Likelihood	1385.27019
AIC (smaller is better)	1387.27019
AICC (smaller is better)	1387.27419
SBC (smaller is better)	1392.17794

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.064022	0.063278	1.0237	0.3117

### Forward Selection: Step 1

Effect Entered: x2

Best 10 Entry Candidates		
Rank	Effect	SBC
1	x2	1370.7425
2	x8	1377.1766
3	x10	1393.3327
4	x3	1397.3647
5	x5	1398.2532
6	x6	1398.5991
7	x4	1398.6078
8	x9	1398.7225
9	x7	1398.9572
10	x1	1399.0848

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

Dimensions	
Columns in Design	2
Number of Effects	2
Max Effect Columns	1
Rank of Design	2

Dimensions	
Parameters in Optimization	2

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	1	28.6037	<.0001

Fit Statistics	
-2 Log Likelihood	1356.66647
AIC (smaller is better)	1360.66647
AICC (smaller is better)	1360.67850
SBC (smaller is better)	1370.48198

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.500671	0.124439	16.1880	<.0001
x2	1	-1.167203	0.221219	27.8386	<.0001

## Forward Selection: Step 2

Effect Entered: x8

Entry Candidates		
Rank	Effect	SBC
1	x8	1356.8279
2	x10	1371.5185
3	x5	1375.5402
4	x3	1375.9472
5	x6	1376.8625
6	x4	1376.9433
7	x9	1376.9659
8	x7	1377.2274
9	x1	1377.3851
10	C	1424.6438

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

Dimensions
------------

Dimensions	
Columns in Design	3
Number of Effects	3
Max Effect Columns	1
Rank of Design	3
Parameters in Optimization	3

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	2	49.2731	<.0001

Fit Statistics	
-2 Log Likelihood	1335.99712
AIC (smaller is better)	1341.99712
AICC (smaller is better)	1342.02122
SBC (smaller is better)	1356.72038

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	1.004603	0.169966	34.9353	<.0001
x2	1	-1.151380	0.223567	26.5231	<.0001
x8	1	-1.030894	0.228869	20.2887	<.0001

## Forward Selection: Step 3

### Effect Entered: x10

Entry Candidates		
Rank	Effect	SBC
1	x10	1358.3616
2	x3	1361.8567
3	x5	1361.8575
4	x9	1363.2118
5	x7	1363.2442
6	x4	1363.3227
7	x6	1363.4113
8	x1	1363.6236

Entry Candidates		
Rank	Effect	SBC
9	C	1409.7123

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

Dimensions	
Columns in Design	4
Number of Effects	4
Max Effect Columns	1
Rank of Design	4
Parameters in Optimization	4

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	3	54.5533	<.0001

Fit Statistics	
-2 Log Likelihood	1330.71689
AIC (smaller is better)	1338.71689
AICC (smaller is better)	1338.75709
SBC (smaller is better)	1358.34791

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.784663	0.194464	16.2812	<.0001
x2	1	-1.157751	0.224308	26.6404	<.0001
x8	1	-1.018780	0.229491	19.7074	<.0001
x10	1	0.075169	0.032835	5.2408	0.0221

## Forward Selection: Step 4

Effect Entered: x3

Entry Candidates		
Rank	Effect	SBC
1	x3	1363.5113
2	x5	1363.6541

Entry Candidates		
Rank	Effect	SBC
3	x7	1364.8625
4	x9	1364.8968
5	x6	1364.9508
6	x4	1365.0212
7	x1	1365.2556
8	C	1411.3321

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

Dimensions	
Columns in Design	5
Number of Effects	5
Max Effect Columns	1
Rank of Design	5
Parameters in Optimization	5

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	4	56.3008	<.0001

Fit Statistics	
-2 Log Likelihood	1328.96942
AIC (smaller is better)	1338.96942
AICC (smaller is better)	1339.02978
SBC (smaller is better)	1363.50820

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.617050	0.231782	7.0873	0.0078
x2	1	-1.153204	0.224562	26.3718	<.0001
x3	1	0.085143	0.064520	1.7415	0.1870
x8	1	-1.028170	0.229826	20.0139	<.0001
x10	1	0.075024	0.032855	5.2143	0.0224



## Effect Entered: x5

Entry Candidates		
Rank	Effect	SBC
1	x5	1368.8719
2	x7	1370.0610
3	x9	1370.1250
4	x6	1370.1265
5	x4	1370.2100
6	x1	1370.4142
7	C	1416.6375

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

Dimensions	
Columns in Design	6
Number of Effects	6
Max Effect Columns	1
Rank of Design	6
Parameters in Optimization	6

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	5	57.8459	<.0001

Fit Statistics	
-2 Log Likelihood	1327.42429
AIC (smaller is better)	1339.42429
AICC (smaller is better)	1339.50888
SBC (smaller is better)	1368.87082

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.494285	0.251895	3.8505	0.0497
x2	1	-1.178319	0.225848	27.2203	<.0001
x3	1	0.083813	0.064584	1.6841	0.1944
x5	1	0.002807	0.002260	1.5425	0.2142

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
x8	1	-1.027560	0.230033	19.9542	<.0001
x10	1	0.073986	0.032894	5.0591	0.0245

### Selection Details

Selection Summary			
Step	Effect Entered	Number Effects In	SBC
0	Intercept	1	1392.1779
1	x2	2	1370.7425
2	x8	3	1356.8279*
3	x10	4	1358.3616
4	x3	5	1363.5113
5	x5	6	1368.8719
* Optimal Value Of Criterion			

Selection stopped at a local minimum of the STOP criterion.

The model at step 2 is selected.

**Selected Effects:** Intercept x2 x8

### Selected Model

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

Dimensions	
Columns in Design	3
Number of Effects	3
Max Effect Columns	1
Rank of Design	3
Parameters in Optimization	3

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq

Testing Global Null Hypothesis: BETA=0			
Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	2	49.2731	<.0001

Fit Statistics	
-2 Log Likelihood	1335.99712
AIC (smaller is better)	1341.99712
AICC (smaller is better)	1342.02122
SBC (smaller is better)	1356.72038

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	1.004603	0.169966	34.9353	<.0001
x2	1	-1.151380	0.223567	26.5231	<.0001
x8	1	-1.030894	0.228869	20.2887	<.0001

Task Timing		
Task	Seconds	Percent
Setup and Parsing	0.01	6.18%
Levelization	0.00	1.68%
Model Initialization	0.00	1.04%
SSCP Computation	0.01	9.30%
Model Selection	0.08	81.56%
Display	0.00	0.10%
Cleanup	0.00	0.00%
Total	0.10	100.00%

In [ ]: