Documentation for FSAA_v12c

DISCLAIMER

All SAS code included in this document is provided "as is" by Bruce Lund without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Recipients acknowledge and agree that Bruce Lund shall not be liable for any damages whatsoever arising out of their use of this material. In addition, Bruce Lund will provide no support for the materials contained herein.

MACRO %FSAA

FSAA fits a logistic regression model by forward selection where the selected variable gives the minimum “adjusted” AIC. The “adjustment” applies to weight of evidence coded variables “C_woe” where the assigned degrees of freedom is computed as a result of a comparison to the alternative entry of the classification variable C into the model at the same step.

There are three stopping criteria for forward selection: (1) stopping at the first local minimum of AIC, (2) after entering all predictors, (3) after a user-specified number of selection steps have been completed.

It is assumed that the logistic model does not exhibit complete of quasi complete separation. Missing values in predictor variables or in the target are handled by PROC LOGISTIC (HPLOGISTIC) in the normal manner.

For details, see the paper “Weight of Evidence, Dummy Variables, and Degrees Of Freedom”, among the papers of the 2021 SAS Global Forum.

DATA SET FOR EXAMPLES

A data set called getStarted in used in this documentation. This dataset is found in SAS documentation by following this link:

https://support.sas.com/documentation/cdl/en/stathpug/66410/HTML/default/viewer.htm#stathpug_hplogistic_gettingstarted01.htm

This dataset has 100 observations and binary target Y. There is a character variable C and 10 numeric variables X1-X10. Other variables (Y_Random and D) are added for the purpose of testing the macro error trapping and processing. In a second DATA Step there is additional processing to collapse C to 7 levels and to create a weight of evidence transform for (collapsed) C and a weight of evidence transform for X2.

```
DATA getStarted;
length C $17;
Y_Random = floor(4*ranuni(1)); /* for testing error conditions */
D = floor(4*ranuni(1));        /* for testing error conditions */
  input C$ Y X1-X10;
  datalines;
  D  0  10.2  6  1.6  38  15  2.4  20  0.8  8.5  3.9
  F  1  12.2  6  2.6  42  61  1.5  10  0.6  8.5  0.7
  D  1   7.7  1  2.1  38  61    1  90  0.6  7.5  5.2
  J  1  10.9  7  3.5  46  42  0.3   0  0.2    6  3.6
  E  0  17.3  6  3.8  26  47  0.9  10  0.4  1.5  4.7
  A  0  18.7  4  1.8   2  34  1.7  80    1  9.5  2.2
  B  0   7.2  1  0.3  48  61  1.1  10  0.8  3.5    4
  D  0   0.1  3  2.4   0  65  1.6  70  0.8  3.5  0.7
  H  1   2.4  4  0.7  38  22  0.2  20    0    3  4.2
  J  0  15.6  7  1.4   0  98  0.3   0    1    5  5.2
  J  0  11.1  3  2.4  42  55  2.2  60  0.6  4.5  0.7
  F  0     4  6  0.9   4  36  2.1  30  0.8    9  4.6
  A  0   6.2  2  1.8  14  79  1.1  70  0.2    0  5.1
  H  0   3.7  3  0.8  12  66  1.3  40  0.4  0.5  3.3
  A  1   9.2  3  2.3  48  51  2.3  50    0    6  5.4
  G  0    14  3    2  18  12  2.2   0    0    3  3.4
  E  1  19.5  6  3.7  26  81  0.1  30  0.6    5  4.8
  C  0    11  3  2.8  38   9  1.7  50  0.8  6.5  0.9
  I  0  15.3  7  2.2  20  98  2.7 100  0.4    7  0.8
```

H	1	7.4	4	0.5	28	65	1.3	60	0.2	9.5	5.4
F	0	11.4	2	1.4	42	12	2.4	10	0.4	1	4.5
C	1	19.4	1	0.4	42	4	2.4	10	0	6.5	0.1
G	0	5.9	4	2.6	12	57	0.8	50	0.4	2	5.8
G	1	15.8	6	3.7	34	8	1.3	90	0.6	2.5	5.7
I	0	10	3	1.9	16	80	3	90	0.4	9.5	1.9
E	0	15.7	1	2.7	32	25	1.7	20	0.2	8.5	6
G	0	11	5	2.9	48	53	0.1	50	1	3.5	1.2
J	1	16.8	0	0.9	14	86	1.4	40	0.8	9	5
D	1	11	4	3.2	48	63	2.8	90	0.6	0	2.2
J	1	4.8	7	3.6	24	1	2.2	20	1	8.5	0.5
J	1	10.4	5	2	42	56	1	20	0	3.5	4.2
G	0	12.7	7	3.6	8	56	2.1	70	1	4.5	1.5
G	0	6.8	1	3.2	30	27	0.6	0	0.8	2	5.6
E	0	8.8	0	3.2	2	67	0.7	10	0.4	1	5
I	1	0.2	0	2.9	10	41	2.3	60	0.2	9	0.3
J	1	4.6	7	3.9	50	61	2.1	50	0.4	3	4.9
J	1	2.3	2	3.2	36	98	0.1	40	0.6	4.5	4.3
I	0	10.8	3	2.7	28	58	0.8	80	0.8	3	6
B	0	9.3	2	3.3	44	44	0.3	50	0.8	5.5	0.4
F	0	9.2	6	0.6	4	64	0.1	0	0.6	4.5	3.9
D	0	7.4	0	2.9	14	0	0.2	30	0.8	7.5	4.5
G	0	18.3	3	3.1	8	60	0.3	60	0.2	7	1.9
F	0	5.3	4	0.2	48	63	2.3	80	0.2	8	5.2
C	0	2.6	5	2.2	24	4	1.3	20	0	2	1.4
F	0	13.8	4	3.6	4	7	1.1	10	0.4	3.5	1.9
B	1	12.4	6	1.7	30	44	1.1	60	0.2	6	1.5
I	0	1.3	1	1.3	8	53	1.1	70	0.6	7	0.8
F	0	18.2	7	1.7	26	92	2.2	30	1	8.5	4.8
J	0	5.2	2	2.2	18	12	1.4	90	0.8	4	4.9
G	1	9.4	2	0.8	22	86	0.4	30	0.4	1	5.9
J	1	10.4	2	1.7	26	31	2.4	10	0.2	7	1.6
J	0	13	1	1.8	14	11	2.3	50	0.6	5.5	2.6
A	0	17.9	4	3.1	46	58	2.6	90	0.6	1.5	3.2
D	1	19.4	6	3	20	50	2.8	100	0.2	9	1.2
I	0	19.6	3	3.6	22	19	1.2	0	0.6	5	4.1
I	1	6	2	1.5	30	30	2.2	20	0.4	8.5	5.3
G	0	13.8	1	2.7	0	52	2.4	20	0.8	6	2
B	0	14.3	4	2.9	30	11	0.6	90	0.6	0.5	4.9
E	0	15.6	0	0.4	38	79	0.4	80	0.4	1	3.3
D	0	14	2	1	22	61	3	90	0.6	2	0.1
C	1	9.4	5	0.4	12	53	1.7	40	0	3	1.1
H	0	13.2	1	1.6	40	15	0.7	40	0.2	9	5.5
A	0	13.5	5	2.4	18	89	1.6	20	0.4	9.5	4.7
E	0	2.6	4	2.3	38	6	0.8	20	0.4	5	5.3
E	0	12.4	3	1.3	26	8	2.8	10	0.8	6	5.8
D	0	7.6	2	0.9	44	89	1.3	50	0.8	6	0.4
I	0	12.7	1	2.3	42	6	2.4	10	0.4	1	3
C	1	10.7	4	3.2	28	23	2.2	90	0.8	5.5	2.8
H	0	10.1	2	2.3	10	62	0.9	50	0.4	2.5	3.7
C	1	16.6	1	0.5	12	88	0.1	20	0.6	5.5	1.8
I	1	0.2	3	2.2	8	71	1.7	80	0.4	0.5	5.5
C	0	10.8	4	3.5	30	70	2.3	60	0.4	4.5	5.9
F	0	7.1	4	3	14	63	2.4	70	0	7	3.1
D	0	16.5	1	3.3	30	80	1.6	40	0	3.5	2.7
H	0	17.1	7	2.1	30	45	1.5	60	0.6	0.5	2.8
D	0	4.3	1	1.5	24	44	0	70	0	5	0.5
H	0	15	2	0.2	14	87	1.8	50	0	4.5	4.7
G	0	19.7	3	1.9	36	99	1.5	10	0.6	3	1.7
H	1	2.8	6	0.6	34	21	2	60	1	9	4.7
G	0	16.6	3	3.3	46	1	1.4	70	0.6	1.5	5.3
E	0	11.7	5	2.7	48	4	0.9	60	0.8	4.5	1.6
F	0	15.6	3	0.2	4	79	0.5	0	0.8	1.5	2.9
C	1	5.3	6	1.4	8	64	2	80	0.4	9	4.2
B	1	8.1	7	1.7	40	36	1.4	60	0.6	6	3.9
I	0	14.8	2	3.2	8	37	0.4	10	0	4.5	3
D	0	7.4	4	3	12	3	0.6	60	0.6	7	0.7
D	0	4.8	3	2.3	44	41	1.9	60	0.2	3	3.1
A	0	4.5	0	0.2	4	48	1.7	80	0.8	9	4.2
D	0	6.9	6	3.3	14	92	0.5	40	0.4	7.5	5
B	0	4.7	4	0.9	14	99	2.4	80	1	0.5	0.7

```
  I  1   7.5  4  2.1  20  79  0.4  40  0.4  2.5  0.7
  C  0   6.1  0  1.4  38  18  2.3  60  0.8  4.5  0.7
  C  0  18.3  1    1  26  98  2.7  20    1  8.5  0.5
  F  0  16.4  7  1.2  32  94  2.9  40  0.4  5.5  2.1
  I  0   9.4  2  2.3  32  42  0.2  70  0.4  8.5  0.3
  F  1  17.9  4  1.3  32  42    2  40  0.2    1  5.4
  H  0  14.9  3  1.6  36  74  2.6  60  0.2    1  2.3
  C  0  12.7  0  2.6   0  88  1.1  80  0.8  0.5  2.1
  F  0   5.4  4  1.5   2   1  1.8  70  0.4  5.5  3.6
  J  1  12.1  4  1.8  20  59  1.3  60  0.4    3  3.8
;
DATA getStarted; set getStarted;
Y_woe = Y; /* for testing error conditions */
if X2 =0 then X2_woe =-0.299242894852857;
if X2 =1 then X2_woe =-0.498769772073583;
if X2 =2 then X2_woe =-0.116231444964965;
if X2 =3 then X2_woe =-1.21507375088538;
if X2 =4 then X2_woe =0.261169230355324;
if X2 =5 then X2_woe =0.106222213255308;
if X2 =6 then X2_woe =1.13606308985493;
if X2 =7 then X2_woe =0.393904285707088;
*;
if C in ( "A","F" ) then C_woe = -0.809318612 ;
if C in ( "B","I","H" ) then C_woe = 0.1069721196 ;
if C in ( "C" ) then C_woe = 0.6177977433 ;
if C in ( "D" ) then C_woe = -0.403853504 ;
if C in ( "E" ) then C_woe = -1.145790849 ;
if C in ( "G" ) then C_woe = -0.703958097 ;
if C in ( "J" ) then C_woe = 1.4932664807 ;
/* collapse levels of C */
if C in ( "A","F") then C = "A_F" ;
if C in ( "B","I","H" ) then C = "B_I_H" ;
run;
```

THE MACRO CALL

Here are the parameters for the macro %FSAA:

```
%FSAA(
  DATASET = , /* data set which contains the predictors and target */
  STEPS = , /* integer >= 1 */
  STOP = , /* FIRST_MIN | space */
  TARGET = , /* binary variable for PROC (HP)LOGISTIC */
  PENALTY = , /* BIC | any other ... defaults to AIC */
  MODEL_DF = , /* 1 if intercept only, else consider &include_var */
  INCLUDE_NUM = , /* numeric var's in model at start of Forward | space */
  INCLUDE_CLASS = , /* class var's in model at start of Forward | space */
  NUM_VAR = , /* numeric var's as candidates to enter */
  CLASS_VAR = , /* classification var's (no suffix "_woe") either NUM or CHAR */
  USE_WOE = , /* woe from class_var's by adding suffix "_woe", e.g. if C_woe is in
                 &USE_WOE, then C must be in &CLASS_VAR */
  ALPHA = ,  /* numeric from open interval (0,1) */
  VERBOSE = , /* YES for more detail | any other is NO */
  HP = , /* HP for HPLOGISTIC | any other ... defaults to LOGISTIC */
  RUN_TITLE = /* Title for run, NO commas allowed */
);
/* Parameters required: DATASET, STEPS, TARGET, MODEL_DF, ALPHA */
/* Parameters NUM_VAR and CLASS_VAR cannot both be spaces */
```

Parameter Documentation:

DATASET names the SAS dataset that contains the dataset to be processed by the macro.

STEPS specifies the maximum number of predictors to be entered in the FORWARD process. This excludes all predictors that are included from **INCLUDE_NUM** and **INCLUDE_CLASS**. Predictors in **INCLUDE_NUM** and **INCLUDE_CLASS** are included in the model before FORWARD process begins.

If STOP = FIRST_MIN then FORWARD process stops when the first (possibly local) minimum for AIC is reached. At least two steps are required. Otherwise, STEPS is the stopping criterion.

TARGET gives the target variable for the logistic model. It must have exactly 2 non-missing levels. It may be character or numeric.

PENALTY defaults to AIC unless PENALTY= BIC.

MODEL_DF gives the degrees of freedom in the model at the start of FORWARD. It is the user's responsibility to set the value for MODEL_DF. If there are no "INCLUDE" predictors, then the user should set MODEL_DF = 1 to account for the intercept. MODEL_DF plays no role in FORWARD but the AIC value of the model will be correct and can be compared to the AIC of other models.

INCLUDE_NUM gives a list of numeric predictors that are included in the model before FORWARD begins. The list may be space.

INCLUDE_CLASS gives a list of classification predictors that are included in the model before FORWARD begins. The list may be space. The list can contain numeric and character predictors. Predictors that appear in INCLUDE_CLASS are placed in the model's CLASS statement.

NUM_VAR lists the numeric variables that are candidates for forward selection.

CLASS_VAR and USE_WOE These two parameters are related. They specify classification variables and weight of evidence variables that are candidates for forward selection.

This usage of these parameters might be confusing to the user and this explanation as well as the examples that follow in a later section should be studied.

Classification predictors (to appear in a CLASS statement) are listed in CLASS_VAR. Weight of evidence predictors appear in USE_WOE with two restrictions: (1) These variables must have the name of the form C_woe (upper or lower case) where C appears in CLASS_VAR. If C does not appear in CLASS_VAR, then this causes an error message and the macro stops. [C is used here to stand for any name for a classification variable, but the name must not exceed 28 characters.]

This entry is allowable:

```
CLASS_VAR C1 C2,
USE_WOE C2_woe,
```

C1 may be selected as a classification variable and C2_woe may be selected as a WOE variable. But C2 is not considered as a classification variable.

This entry is not allowable:

```
CLASS_VAR C1,
USE_WOE C2_woe,
```

This entry is allowable (but does not utilize the functionality of the macro … to adjust d.f. for WOE's):

```
CLASS_VAR C1 C2,
USE_WOE ,
```

ALPHA This is used in determining d.f. for entering a weight of evidence variable. See the paper "Weight of Evidence, Dummy Variables, and Degrees Of Freedom". ALPHA is required and must be a number in the open interval (0, 1).

VERBOSE If YES, then step by step details are printed.

HP If HP, then PROC HPLOGISTIC is used by the macro, otherwise PROC LOGISTIC. The results of the macro processing are unchanged apart from computational differences between these procedures.

RUN_TITLE An optional title1 for the macro reporting. Otherwise, "No Title" appears.

ERROR CHECKING:

These input and parameter errors are detected by the %FSAA macro.

```
1.  Check for missing required parameters DATASET, STEPS, TARGET, MODEL_DF, ALPHA
2.  Check that there is no "dash" in parameters which list variables
   (TARGET INCLUDE_NUM INCLUDE_CLASS NUM_VAR CLASS_VAR USE_WOE)
3.  Check that TARGET has only one variable
4.  Check that not both NUM_VAR and CLASS_VAR are space
5.  Check that DATASET exists
6.  Check that DATASET has observations
7.  Check that ALPHA is numeric in (0, 1)
8.  Check that STEPS is a positive integer
9.  Check that MODEL_DF is a positive integer
10. Check that variables in "input parameters" exist in DATASET
   (TARGET INCLUDE_NUM INCLUDE_CLASS NUM_VAR CLASS_VAR USE_WOE)
11. Check that all var in "USE_WOE" have suffix "_woe"
12. Check that INCLUDE_NUM and NUM_VAR have numeric variables
13. Check that there are no duplicates among input variables
   (TARGET INCLUDE_NUM INCLUDE_CLASS NUM_VAR CLASS_VAR USE_WOE)
14. Check that variables in USE_WOE has associated variable in CLASS_VAR
15. Check that TARGET has 2 levels and that no variable in INCLUDE_NUM INCLUDE_CLASS
    NUM_VAR CLASS_VAR USE_WOE is all missing or has only 1 non-missing level
```

SOME OF THE ERROR CHECKS ARE ILLUSTRATED BELOW:

```
2.  Check that there is no "dash" in parameters which give a list of variables
   (TARGET INCLUDE_NUM INCLUDE_CLASS NUM_VAR CLASS_VAR USE_WOE)
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, penalty = , model_df = 1,
include_num = , include_class = , num_var = X1 - X4, class_var = C,
use_woe = C_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #2
);
```

Obs	error_msg
1	dash convention is not supported for input variable lists
2	check: target, include_num, include_class, num_var, class_var, use_woe
3	2. ENDING EXECUTION

```
3.  Check that TARGET has only one variable
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y Y_Random, model_df = 1,
include_num = , include_class = , num_var = X1 X2, class_var = C,
use_woe = C_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #3
);
```

Obs	error_msg
1	target contains multiple components separate by spaces
2	target = Y Y_Random
3	3. ENDING EXECUTION

```
4.  Check that not both NUM_VAR and CLASS_VAR are space
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, model_df = 1,
include_num = X1, include_class = C, num_var = , class_var = ,
use_woe = C_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #4
);
```

Obs	error_msg
1	Not both num_var and class_var can be SPACE
2	At least one of these parameters must be non-space
3	4. ENDING EXECUTION

```
10. Check that variables in "input parameters" exist in DATASET
   (TARGET INCLUDE_NUM INCLUDE_CLASS NUM_VAR CLASS_VAR USE_WOE)
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, model_df = 1,
include_num = X1, include_class = C, num_var = X, class_var = ,
use_woe = C_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #10
);
```

Obs	error_msg
1	One or more input variables are not in getStarted
2	CHECK: target include_num include_class num_var class_var use_woe
3	10. ENDING EXECUTION

```
11. Check that all variables in "USE_WOE" have suffix "_woe"
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, model_df = 1,
include_num = X1, include_class = C, num_var = X1, class_var = C X2,
use_woe = C, alpha = .05, verbose = , HP = , RUN_TITLE = Error #11
);
```

Obs	error_msg
1	A woe variable does not have suffix _woe
2	C
3	11. ENDING EXECUTION

Comment: Only one violation (i.e. C) is reported even if there is more than one.

```
12. Check that INCLUDE_NUM and NUM_VAR have numeric variables
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, model_df = 1,
include_num = C, include_class = , num_var = X1 X3, class_var = X2,
use_woe = X2_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #12
);
```

Obs	error_msg
1	One or more input variables is not numeric as required
2	CHECK: include_num num_var
3	12. ENDING EXECUTION

```
13. Check that there are no duplicates among input variables
   (TARGET INCLUDE_NUM INCLUDE_CLASS NUM_VAR CLASS_VAR USE_WOE)
```

```
%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, model_df = 1,
include_num = X1, include_class = X2, num_var = X1 X3, class_var = X2,
use_woe = X2_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #13
);
```

Obs	Duplicated
1	X2
2	X1

Obs	error_msg
1	Duplicate in Predictors and/or Target
2	13. ENDING EXECUTION

```
14. Check that variables in USE_WOE have associated variable in CLASS_VAR

%FSAA(
dataset = getStarted, steps = 6, stop = , target = Y, model_df = 1,
include_num = , include_class = , num_var = X1 X3, class_var = ,
use_woe = X2_woe C_woe, alpha = .05, verbose = , HP = , RUN_TITLE = Error #14
);
```

Obs	error_msg
1	woe variable in use_woe without variable in class_var
2	C_woe
3	14. ENDING EXECUTION

Comment: Only one violation (i.e. C_woe) is reported even if there is more than one.

Example #1

Predictor C may enter as a classification variable, X2_woe may enter as weight of evidence, and X1 is included in the model. The model_df is set to 2 (intercept, X1). There are 3 candidate numeric predictors X3 X4 X5.

Stop = first_min (if a local minimum AIC is found, then the selection process stops).
Steps = 4 (at most 4 candidate predictors will be entered).
Alpha = 0.05
Verbose and HP have default values.
Target = Y

```
%FSAA(
dataset = getStarted, steps = 4, stop = first_min, target = Y, penalty = ,
model_df = 2, include_num = X1, include_class = , num_var = X4 X3 X5,
class_var = C X2, use_woe = X2_woe, alpha = .05, verbose = , HP = ,
RUN_TITLE = FSAA Model #1
);
```

FSAA Model #1
FSAA Version v10, RUN ON 08-NOV-2020 13:35:33
Dataset= getStarted, Target= Y
Included Variables at Start

Obs	included_variables	model_df
1	X1	2

FSAA Model #1
FSAA Version v10, RUN ON 08-NOV-2020 13:35:33
Dataset= getStarted, Target= Y
Count of Levels in CLASS Variables

Obs	class_variable	class_levels
1	X2	8
2	C	7

FSAA Model #1
FSAA Version v10, RUN ON 08-NOV-2020 13:35:33
Dataset= getStarted, Target= Y

Obs	msg
1	#3 Local minimum AIC reached
2	At Step = 1
3	Exit Loop

FSAA Model #1
FSAA Version v10, RUN ON 08-NOV-2020 13:35:33
Dataset= getStarted, Target= Y
Summary Report

Obs	step	min AIC var	min adj AIC	best model	adj-df for min	new model df	new included var
1	1	X4	124.485	*	1	3	X1 X4
2	2	C	125.066		6	9	X1 X4 C

FSAA Model #1
FSAA Version v10, RUN ON 08-NOV-2020 13:35:33
Dataset= getStarted, Target= Y
Predictors not in the model

Obs	step	num_var	class_var
1	1	X3 X5	X2 C
2	2	X3 X5	X2

Example #2

Predictor C_woe may enter as weight of evidence. X1 is included in the model. The model_df is set to 2 (intercept, X1). There are 3 candidate numeric predictors X2 X8 X10

Stop = (Ignores local minimum AIC).
Steps = 4 (at most 4 candidate predictors will be entered).
Alpha = 0.05
Verbose and HP have default values.
Target = Y

```
%FSAA(
dataset = getStarted, steps = 4, stop = , target = Y, model_df = 2,
include_num = X1, include_class = , num_var = X2 X8 X10, class_var = C,
use_woe = C_woe, alpha = .05, verbose = , HP = , RUN_TITLE = FSAA Model #2
);
```

(Partial output)

FSAA Model #2
FSAA Version v10, RUN ON 08-NOV-2020 13:48:21
Dataset= getStarted, Target= Y

Obs	msg
1	#2 All predictors entered
2	At STEP = 4
3	Exit Loop

FSAA Model #2
FSAA Version v10, RUN ON 08-NOV-2020 13:48:21
Dataset= getStarted, Target= Y
Summary Report

Obs	step	min AIC var	min adj AIC	best model	adj-df for min	new model df	new included var
1	1	X2	124.116		1	3	X1 X2
2	2	X8	120.746		1	4	X1 X2 X8
3	3	C_woe	119.983	*	6	10	X1 X2 X8 C_woe
4	4	X10	120.734		1	11	X1 X2 X8 C_woe X10

Example #3

Predictor C_woe, X2_woe may enter as weight of evidence. X1 is included in the model. The model_df is set to 2 (intercept, X1). There are 0 candidate numeric predictors

Stop = (Ignores local minimum AIC).
Steps = 4 (at most 4 candidate predictors will be entered … but only two predictors are candidates).
Alpha = 0.25
Verbose and HP have default values.
Target = Y

```
%FSAA(
dataset = getStarted, steps = 4, stop = , target = Y, model_df = 2,
include_num = X1, include_class = , num_var = , class_var = X2 C,
use_woe = X2_woe C_woe, alpha = .25, verbose = , HP = , RUN_TITLE = FSAA Model #3
);
```

(Partial output)

FSAA Model #3
FSAA Version v10, RUN ON 08-NOV-2020 13:52:09
Dataset= getStarted, Target= Y

Obs	msg
1	#2 All predictors entered
2	At STEP = 2
3	Exit Loop

Notice there are fractional d.f. assigned to C_woe and X2_woe.

FSAA Model #3
FSAA Version v10, RUN ON 08-NOV-2020 13:52:09
Dataset= getStarted, Target= Y
Summary Report

Obs	step	min AIC var	min adj AIC	best model	adj-df for min	new model df	new included var
1	1	C_woe	126.659		5.8	7.8	X1 C_woe
2	2	X2_woe	126.213	*	5.6	13.4	X1 C_woe X2_woe

Example #4

All candidate predictors are numeric. Since there are no WOE predictors, the FSAA macro should not be used (inefficient). There are no adjustments to the d.f. of candidate predictors.

```
%FSAA(
dataset = getStarted, steps = 4, stop = , target = Y, model_df = 2,
include_num = X1, include_class = , num_var = X3 X4 X5 X6 X7 X8 X9, class_var = ,
use_woe = , alpha = .25, verbose = , HP = , RUN_TITLE = FSAA Model #4
);
```

(Partial output)

FSAA Model #4
FSAA Version v10, RUN ON 08-NOV-2020 13:59:21
Dataset= getStarted, Target= Y

Obs	msg
1	#1 All Steps Completed
2	STEP = 4
3	End of Loop

FSAA Model #4
FSAA Version v10, RUN ON 08-NOV-2020 13:59:21
Dataset= getStarted, Target= Y
Summary Report

Obs	step	min AIC var	min adj AIC	best model	adj-df for min	new model df	new included var
1	1	X8	124.281		1	3	X1 X8
2	2	X4	122.545		1	4	X1 X8 X4
3	3	X9	121.538	*	1	5	X1 X8 X4 X9
4	4	X5	122.549		1	6	X1 X8 X4 X9 X5

Example #5

Example #5 shows all FSAA reports with the exception of reports produced by VERBOSE = YES.

In dataset TEST the classification predictors C1 - C5 play no role in the equation for the target Y. These C-C5 are converted to weight of evidence in the second DATA Step, C1_woe, C2_woe, C3_woe, C4_woe, C5_woe. Only C4_woe enters the model. This example is presented in my SGF 2021 video.

(The latest version, v12c, was used to run Example #5. The titles and formatting reflects this latest version.)

```
data test;
do i = 1 to 5000;
       call streaminit(12345);
       C1 = floor(10*rand('uniform'));
       C2 = floor(9*rand('uniform'));
       C3 = floor(7*rand('uniform'));
       C4 = floor(5*rand('uniform'));
       C5 = floor(3*rand('uniform'));
       B1 = 2*floor(2*rand('uniform')) - 1;
       B2 = 2*floor(2*rand('uniform')) - 1;
       N1 = rand('normal');
       N2 = rand('normal');
       U = rand('uniform'); e = log(U/(1-U));
       Y_star = 1*B1 + 2*B2 + 1*N1 + 2*N2 + e;
       Y = (Y_star > 0);
       output;
       end;
run;
data test2; set test;
if C1 in ( 0 ) then C1_woe = -0.120296147 ;
if C1 in ( 1 ) then C1_woe = -0.0914636 ;
if C1 in ( 2 ) then C1_woe = -0.033780334 ;
if C1 in ( 3 ) then C1_woe = 0.0936619553 ;
if C1 in ( 4 ) then C1_woe = 0.0760205948 ;
if C1 in ( 5 ) then C1_woe = 0.1009817326 ;
if C1 in ( 6 ) then C1_woe = -0.033612409 ;
if C1 in ( 7 ) then C1_woe = -0.008736447 ;
if C1 in ( 8 ) then C1_woe = 0.1291250313 ;
if C1 in ( 9 ) then C1_woe = -0.104426927 ;
if C2 in ( 0 ) then C2_woe = 0.0205938085 ;
if C2 in ( 1 ) then C2_woe = 0.0588092989 ;
if C2 in ( 2 ) then C2_woe = 0.0768059205 ;
if C2 in ( 3 ) then C2_woe = -0.098884862 ;
if C2 in ( 4 ) then C2_woe = -0.070039706 ;
if C2 in ( 5 ) then C2_woe = 0.089776008 ;
if C2 in ( 6 ) then C2_woe = 0.0542031185 ;
if C2 in ( 7 ) then C2_woe = -0.066076105 ;
if C2 in ( 8 ) then C2_woe = -0.058013302 ;
if C3 in ( 0 ) then C3_woe = -0.043254163 ;
```

```
if C3 in ( 1 ) then C3_woe = 0.0286067978 ;
if C3 in ( 2 ) then C3_woe = 0.0181202911 ;
if C3 in ( 3 ) then C3_woe = -0.013934687 ;
if C3 in ( 4 ) then C3_woe = -0.017745622 ;
if C3 in ( 5 ) then C3_woe = -0.010205457 ;
if C3 in ( 6 ) then C3_woe = 0.0353445689 ;
if C4 in ( 0 ) then C4_woe = -0.083160962 ;
if C4 in ( 1 ) then C4_woe = -0.072055659 ;
if C4 in ( 2 ) then C4_woe = 0.0543469103 ;
if C4 in ( 3 ) then C4_woe = 0.0215844407 ;
if C4 in ( 4 ) then C4_woe = 0.0734005026 ;
if C5 in ( 0 ) then C5_woe = 0.0244679979 ;
if C5 in ( 1 ) then C5_woe = -0.057358159 ;
if C5 in ( 2 ) then C5_woe = 0.0304969945 ;

%FSAA(
dataset = test2,
steps = 12,
stop = ,
target = Y,
penalty = ,
model_df = 1,
include_num = ,
include_class = ,
num_var = B1 B2 N1 N2,
class_var = C1 C2 C3 C4 C5,
use_woe = C1_woe C2_woe C3_woe C4_woe C5_woe,
alpha = .05,
verbose = ,
HP = HP,
RUN_TITLE = FSAA Model #5
);
```

NOTICE

Obs	notice
1	Running FSAA version = v12c
2	This program is provided without warranty.
3	See Disclaimer in FSAA_v12c.sas for important information.
4	WARNING: There is only partial Input Data and Parameter Checking in this Program
5	WARNING: This is a beta-version and remains under development

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y
Number of levels of target and predictors

Obs	TableVar	NLevels	NMissLevels	NNonMissLevels
1	Y	2	0	2
2	B1	2	0	2
3	B2	2	0	2
4	N1	5000	0	5000
5	N2	5000	0	5000
6	C1	10	0	10
7	C2	9	0	9
8	C3	7	0	7
9	C4	5	0	5
10	C5	3	0	3
11	C1_woe	10	0	10

Obs	TableVar	NLevels	NMissLevels	NNonMissLevels
12	C2_woe	9	0	9
13	C3_woe	7	0	7
14	C4_woe	5	0	5
15	C5_woe	3	0	3

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y

Obs	REPORT_MSG
1	Parameter and Data Check was Successful
2	.....options selected:
3	Running PROC HPLOGISTIC
4	Stopping FORWARD at steps= 12 or when all variables are entered
5	Print summary results Only
6	Penalty term: AIC
7	Alpha = .05, Initial model d.f. = 1
8	.....For rows below: only first 130 characters are printed
9	Included numeric:
10	Included class:
11	Numeric predictors: B1 B2 N1 N2
12	Class predictors: C1 C2 C3 C4 C5
13	WOE predictors candidates: C1_woe C2_woe C3_woe C4_woe C5_woe

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y
Included Variables at Start

Obs	included_variables	model_df
1		1

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y
Count of Levels in CLASS Variables

Obs	class_variable	class_levels
1	C1	10
2	C2	9
3	C3	7
4	C4	5
5	C5	3

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y

Obs	msg
1	#2 All predictors entered
2	At STEP = 9
3	Exit Loop

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y
Summary Report
Penalty is AIC
Observations Used= 5000

Obs	step	min AIC var	min adj AIC	best model	adj-df for min	new model df	new included var
1	1	B2	5753.50		1.0	2.0	B2
2	2	N2	4375.79		1.0	3.0	B2 N2
3	3	B1	3960.51		1.0	4.0	B2 N2 B1
4	4	N1	3496.29		1.0	5.0	B2 N2 B1 N1
5	5	C4_woe	3490.01	*	3.8	8.8	B2 N2 B1 N1 C4_woe
6	6	C5_woe	3490.73		2.0	10.8	B2 N2 B1 N1 C4_woe C5_woe
7	7	C2_woe	3498.45		7.7	18.5	B2 N2 B1 N1 C4_woe C5_woe C2_woe
8	8	C3_woe	3508.36		5.1	23.6	B2 N2 B1 N1 C4_woe C5_woe C2_woe C3_woe
9	9	C1_woe	3521.94		7.8	31.4	B2 N2 B1 N1 C4_woe C5_woe C2_woe C3_woe C1_woe

FSAA Model #5
FSAA Version v12c, RUN ON 15-FEB-2021 11:34:10
Dataset= test2, Target= Y
Predictors not in the model

Obs	step	num_var	class_var
1	1	B1 N1 N2	C1 C2 C3 C4 C5
2	2	B1 N1	C1 C2 C3 C4 C5
3	3	N1	C1 C2 C3 C4 C5
4	4		C1 C2 C3 C4 C5
5	5		C1 C2 C3 C5
6	6		C1 C2 C3
7	7		C1 C3
8	8		C1
9	9		

Example #5b

Example #5b modifies Example #5 by treating C1_woe, C2_woe, C3_woe, C4_woe, C5_woe as numeric predictors with 1 d.f. The minimum AIC is reached at Step 8. All predictors except C3_woe are selected. Note that C1_woe, C2_woe, C3_woe, C4_woe, C5_woe played no role into the equation defining the target Y. The result of treating C1_woe, C2_woe, C3_woe, C4_woe, C5_woe as having 1 d.f. is to select four meaningless predictors C1_woe, C2_woe, C4_woe, C5_woe for the model.

This example is presented in my SGF 2021 video.

```
%FSAA(
dataset = test2,
steps = 12,
stop = ,
target = Y,
penalty = ,
model_df = 1,
include_num = ,
include_class = ,
num_var = B1 B2 N1 N2 C1_woe C2_woe C3_woe C4_woe C5_woe,
class_var = ,
use_woe = ,
```

```
alpha = .05,
verbose = ,
HP = HP,
RUN_TITLE = Regard WOE as numeric with 1 d.f.
);
```

Regard WOE as numeric with 1 d.f.
FSAA Version v12c, RUN ON 16-FEB-2021 12:36:55
Dataset= test2, Target= Y
Summary Report
Penalty is AIC
Observations Used= 5000

Obs	step	min AIC var	min adj AIC	best model	adj-df for min	new model df	new included var
1	1	B2	5753.50		1	2	B2
2	2	N2	4375.79		1	3	B2 N2
3	3	B1	3960.51		1	4	B2 N2 B1
4	4	N1	3496.29		1	5	B2 N2 B1 N1
5	5	C4_woe	3484.41		1	6	B2 N2 B1 N1 C4_woe
6	6	C2_woe	3478.67		1	7	B2 N2 B1 N1 C4_woe C2_woe
7	7	C5_woe	3477.45		1	8	B2 N2 B1 N1 C4_woe C2_woe C5_woe
8	8	C1_woe	3477.40	*	1	9	B2 N2 B1 N1 C4_woe C2_woe C5_woe C1_woe
9	9	C3_woe	3479.14		1	10	B2 N2 B1 N1 C4_woe C2_woe C5_woe C1_woe C3_woe