IHS Automotive
DRIVEN BY POLK

Macromatic

A Case Study in Building a Macro

Barbara Keys, Analytic Consultant
October 24, 2013
Presented to Michigan SAS Users Group

The Task

IHS Automotive
DRIVEN BY POLK

- Build a logistic regression on all possible combinations of three independent variables
- Capture the test statistics for each combination from the test of Global Null Hypothesis
  - Likelihood Ratio
  - Score
  - Wald

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

The Data

IHS Automotive
DRIVEN BY POLK

- Almost 29K zip codes
- Series of demographic and aggregated vehicle percentages at a zip code level:
  - Percent of zip code with income > $100K
  - Percent of zip code owning a vehicle bought new
  - 54 in all, although only 10 are used in this presentation
  - Almost 25K unique subsets of three from a list of 54
- Model probability of a household in the zip code buying a new vehicle in a defined time period
- DV is in the form events/trials:
  - Trials = number of households in zip code
  - Events = number of households buying new vehicle in zip code

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

The Plan

IHS Automotive
DRIVEN BY POLK

- Do it once – write a macro-less program to accomplish one run of the task
- Turn this program into a macro (inner macro)
- Create a macro to call the inner macro repeatedly and harvest the results from each run (outer macro)
- Fine tune as needed
- Warning – memory is not your friend!

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

First Run

IHS Automotive
DRIVEN BY POLK

PROC LOGISTIC DATA=STUDY;
MODEL BUYCNT/HHDCNT=XRMI011 XRMI020 XRMI028;
RUN;

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	**250396.743**	3	<.0001
Score	**252064.487**	3	<.0001
Wald	**251579.812**	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.5567	0.00617	171584.146	<.0001
XRMI011	1	0.00198	0.000036	3001.9855	<.0001
XRMI020	1	0.00443	0.000015	90998.4877	<.0001
XRMI028	1	-0.0468	0.000123	143605.532	<.0001

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Using ODS to Find Needed Output

IHS Automotive
DRIVEN BY POLK

ODS TRACE ON;
PROC LOGISTIC DATA=STUDY;
MODEL BUYCNT/HHDCNT=XRMI011 XRMI020 XRMI028;
RUN;
ODS TRACE OFF;

```
 Output Added:
-------------
Name:         GlobalTests
Label:        Global Tests
Template:     Stat.Logistic.GlobalTests
Path:         Logistic.GlobalTests
-------------
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Using ODS to Find Output

IHS Automotive
DRIVEN BY POLK

```
ODS OUTPUT GLOBALTESTS=TESTS;
PROC LOGISTIC DATA=STUDY;
MODEL BUYCNT/HHDCNT=XRMI011 XRMI020 XRMI028;
RUN;
PROC PRINT DATA=TESTS;
RUN;
```

Obs	Test	ChiSq	DF	Prob ChiSq
1	Likelihood Ratio	250396.743	3	<.0001
2	Score	252064.487	3	<.0001
3	Wald	251579.812	3	<.0001

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Using ODS to Capture Output

IHS Automotive
DRIVEN BY POLK

```
DATA TESTS2(KEEP=LIKELIHOOD SCORE WALD); SET TESTS;
RETAIN LIKELIHOOD SCORE WALD;

IF _N_=1 THEN LIKELIHOOD=INPUT(CHISQ,12.3);
ELSE IF _N_=2 THEN SCORE=INPUT(CHISQ,12.3);
ELSE IF _N_=3 THEN DO;
WALD=INPUT(CHISQ,12.3);
OUTPUT;
END;
```

ODS stores the values as character, so need to convert to numeric

LIKELIHOOD	SCORE	WALD
250396.74	252064.49	251579.81

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Inner Macro (First Attempt)

IHS Automotive
DRIVEN BY POLK

```
%MACRO LOGIT(IVAR1,IVAR2,IVAR3);

ODS OUTPUT GLOBALTESTS=TESTS;

PROC LOGISTIC DATA=STUDY;
MODEL BUYCNT/HHDCNT= &IVAR1 &IVAR2 &IVAR3;
RUN;
QUIT;

DATA TESTS2(KEEP=LIKELIHOOD SCORE WALD VAR1 VAR2
VAR3);
SET TESTS;
RETAIN LIKELIHOOD SCORE WALD;

IF _N_=1 THEN LIKELIHOOD=INPUT(CHISQ,12.3);
ELSE IF _N_=2 THEN SCORE=INPUT(CHISQ,12.3);
ELSE IF _N_=3 THEN DO;
WALD=INPUT(CHISQ,12.3);
OUTPUT;
END;
RUN;

%MEND
```

LIKELIHOOD	SCORE	WALD
250396.74	252064.49	251579.81

%LOGIT(XRMI011,XRMI020,XRMI028)

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Inner Macro w/ Input Names

IHS Automotive
DRIVEN BY POLK

```
%MACRO LOGIT(IVAR1,IVAR2,IVAR3);

ODS OUTPUT GLOBALTESTS=TESTS;

PROC LOGISTIC DATA=STUDY;
MODEL BUYCNT/HHDCNT= &IVAR1 &IVAR2 &IVAR3;
RUN;
QUIT;

DATA TESTS2(KEEP=LIKELIHOOD SCORE WALD VAR1 VAR2 VAR3); SET TESTS;
RETAIN LIKELIHOOD SCORE WALD;
VAR1="&IVAR1";
VAR2="&IVAR2";
VAR3="&IVAR3";

IF _N_=1 THEN LIKELIHOOD=INPUT(CHISQ,12.3);
ELSE IF _N_=2 THEN SCORE=INPUT(CHISQ,12.3);
ELSE IF _N_=3 THEN DO;
WALD=INPUT(CHISQ,12.3);
OUTPUT;
END;
RUN;

%MEND
```

VAR1	VAR2	VAR3	LIKELIHOOD	SCORE	WALD
XRMI011	XRMI020	XRMI028	250396.74	252064.49	251579.81

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Create a File of All Possible Combos

IHS Automotive
DRIVEN BY POLK

```
PROC FORMAT;
VALUE VLIST
1  =  'XRMI011'
2  =  'XRMI020'
3  =  'XRMI023'
4  =  'XRMI028'
5  =  'XRMI030'
6  =  'XRMI035'
7  =  'XRMI037'
8  =  'XRMI040'
9  =  'XRMI042'
10  =  'XRMI045';

DATA VCOMBO;
DO I = 1 TO 10;
 DO J = I+1 TO 10;
  DO K = J+1 TO 10;
  VAR1=PUT(I,VLIST.);
  VAR2=PUT(J,VLIST.);
  VAR3=PUT(K,VLIST.);
  KEEP VAR1 VAR2 VAR3;
OUTPUT;
END;
END;
END;
RUN;
```

Format statement to translate consecutive numbers to variable names

Writes out all possible combos to a SAS dataset

Obs	VAR1	VAR2	VAR3
1	XRMI011	XRMI020	XRMI023
2	XRMI011	XRMI020	XRMI028
3	XRMI011	XRMI020	XRMI030
4	XRMI011	XRMI020	XRMI035
5	XRMI011	XRMI020	XRMI037
6	XRMI011	XRMI020	XRMI040
7	XRMI011	XRMI020	XRMI042
8	XRMI011	XRMI020	XRMI045
9	XRMI011	XRMI023	XRMI028
10	XRMI011	XRMI023	XRMI030

```
DATA VCOMBO; SET VCOMBO;
  CALL SYMPUT("NUMCOMB",_N_);
RUN;
```

Calculate how many combos and write the result to a macro variable, needed for an index counter

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Outer Macro

IHS Automotive
DRIVEN BY POLK

Set up files to hold list and log output for PROC PRINTO

```
%MACRO RLOGIT;

FILENAME OUTLIST 'C:\AAAWORK\MISUG\LIST.TXT';
FILENAME OUTLOG 'C:\AAAWORK\MISUG\LOG.TXT';

PROC PRINTTO PRINT=OUTPUT LOG=OUTLOG NEW;
RUN;

%DO I=1 %TO &NUMCOMB;
DATA _NULL_ ; SET VCOMBO;
IF _N_=&I;
CALL SYMPUT("VAR1",VAR1);
CALL SYMPUT("VAR2",VAR2);
CALL SYMPUT("VAR3",VAR3);
RUN;

%LOGIT(&VAR1,&VAR2,&VAR3)
```

Macro DO loop calls the inner macro once for every line in the file of all possible combos

Write output from each call to a RESULTS file

```
 %IF &I=1 %THEN %DO;
DATA RESULTS; SET TESTS2;
RUN;
%END;
%ELSE %DO;
DATA RESULTS; SET RESULTS TESTS2;
RUN;
%END;
%END;
PROC PRINTTO;
RUN;
%MEND;
```

%RLOGIT

VAR1	VAR2	VAR3	LIKELIHOOD	SCORE	WALD
XRMI011	XRMI020	XRMI023	135807.58	134168.84	133533.12
XRMI011	XRMI020	XRMI028	250396.74	252064.49	251579.81
XRMI011	XRMI020	XRMI030	105814.20	103977.50	103262.31
XRMI011	XRMI020	XRMI035	119180.88	117825.01	117179.07
XRMI011	XRMI020	XRMI037	351694.18	377410.89	370517.92

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

It Works, but . . .

IHS Automotive
DRIVEN BY POLK

- You have to write a custom format for each job
- Filenames are hard-coded
- Dependent variable name is hard-coded
- Number of input variable is hard-coded
- Need to get results out of the SAS dataset and put them into a display format

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Tackle Format Statement First

IHS Automotive
DRIVEN BY POLK

```
DATA VARS;
INPUT VNAME $;
SEQUENCE=_N_;
CARDS;
XRMI011
XRMI020
  etc.
;

FILENAME FMT 'C:\AAAWORK\MISUG\VFORMAT.TXT';

DATA _NULL_; SET VARS END=LAST; FILE FMT;
IF _N_=1 THEN DO;
PUT "PROC FORMAT;" ;
PUT "VALUE VLIST";
END;
PUT SEQUENCE "= '" VNAME "'";
IF LAST THEN DO;
PUT ";";
COMBOS=COMB(_N_,3);
CALL SYMPUT("NUMCOMB",COMBOS);
CALL SYMPUT("NUMVAR",_N_);
END;
RUN;

%INCLUDE FMT;
```

Create a SAS dataset containing only names of input variables

Writes a text file containing format statement and then %INCLUDEs it

Stores the number of input variable and the number of resulting combos as macro variables

The created text file:

```
PROC FORMAT;
VALUE VLIST
1 = 'XRMI011 '
2 = 'XRMI020 '
3 = 'XRMI023 '
4 = 'XRMI028 '
5 = 'XRMI030 '
6 = 'XRMI035 '
7 = 'XRMI037 '
8 = 'XRMI040 '
9 = 'XRMI042 '
10 = 'XRMI045 '
;
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Additional Macro Parameters

IHS Automotive
DRIVEN BY POLK

- Pass SAS dataset names and name of dependent variable as positional parameters

```
%MACRO RLOGIT(VARS,SASDS,DV);
```

- VARS = name of SAS dataset containing list of inputs
- SASDS = name of SAS dataset containing model-building data
- DV = name of dependent variable

- Use %LET to pass filenames

```
%LET FMT=C:\AAAWORK\MISUG\VFORMAT.TXT; /* text file for formats */
%LET OUTLIST=C:\AAAWORK\MISUG\LIST.TXT; /* file to capture output window */
%LET OUTLOG=C:\AAAWORK\MISUG\LOG.TXT; /* file to capture log window */
%LET RESULTS=C:\AAAWORK\MISUG\RESULTS.HTML; /* HTML file to hold results */
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

ODS HTML Destination to Display Results

IHS Automotive
DRIVEN BY POLK

- Add ODS commands to outer macro to display results

```
ODS LISTING CLOSE;
ODS HTML BODY="&RESULTS";

PROC PRINT DATA=RESULTS NOOBS;
VAR VAR1 VAR2 VAR3 LIKELIHOOD SCORE WALD;
RUN;

ODS HTML CLOSE;
ODS LISTING;
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Summary of Steps

IHS Automotive
DRIVEN BY POLK

Create SAS dataset of predictors

Build format statement

Create dataset of all possible combos

Select one combo at a time and send to inner macro

- Perform logistic regression
- Capture needed stats

Receive results from inner macro and accumulate them into a dataset

Display final results in an HTML file

Black font = processing outside of macro
Red font = outer macro processing
Blue font = inner macro processing

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Outer Macro

IHS Automotive
DRIVEN BY POLK

```
%MACRO RLOGIT(VARS,SASDS,DV);

FILENAME FMT "&FMT";

DATA _NULL_; SET &VARS END=LAST; FILE FMT;
IF _N_=1 THEN DO;
PUT "PROC FORMAT;" ;
PUT "VALUE VLIST";
END;
PUT SEQUENCE "= '" VNAME "'";
IF LAST THEN DO;
PUT ";";
COMBOS=COMB(_N_,3);
CALL SYMPUT("NUMCOMB",COMBOS);
CALL SYMPUT("NUMVAR",_N_);
END;
RUN;

%INCLUDE FMT;

FILENAME OUTLIST "&OUTLIST";
FILENAME OUTLOG "&OUTLOG";

PROC PRINTTO PRINT=OUTLIST LOG=OUTLOG NEW;
RUN;

DATA VCOMBO;
DO I = 1 TO &NUMVAR;
 DO J = I+1 TO &NUMVAR;
  DO K = J+1 TO &NUMVAR;
  VAR1=PUT(I,VLIST.);
  VAR2=PUT(J,VLIST.);
  VAR3=PUT(K,VLIST.);
  KEEP VAR1 VAR2 VAR3;
OUTPUT;
END;
END;
END;
RUN;

%DO I=1 %TO &NUMCOMB;
DATA _NULL_ ; SET VCOMBO;
IF _N_=&I;
CALL SYMPUT("VAR1",VAR1);
CALL SYMPUT("VAR2",VAR2);
CALL SYMPUT("VAR3",VAR3);
RUN;

%LOGIT(&VAR1,&VAR2,&VAR3)

%IF &I=1 %THEN %DO;
DATA RESULTS; SET TESTS2;
RUN;
%END;
%ELSE %DO;
DATA RESULTS; SET RESULTS TESTS2;
RUN;
%END;
%END;
PROC PRINTTO;
RUN;

ODS LISTING CLOSE;
ODS HTML BODY="&RESULTS";

PROC PRINT DATA=RESULTS NOOBS;
VAR VAR1 VAR2 VAR3 LIKELIHOOD SCORE WALD;
RUN;

ODS HTML CLOSE;
ODS LISTING;
%MEND;
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Inner Macro

IHS Automotive
DRIVEN BY POLK

```
%MACRO LOGIT(IVAR1,IVAR2,IVAR3);

ODS OUTPUT GLOBALTESTS=TESTS;

PROC LOGISTIC DATA=&SASDS;
MODEL &DV= &IVAR1 &IVAR2 &IVAR3;
RUN;
QUIT;

DATA TESTS2(KEEP=LIKELIHOOD SCORE WALD VAR1
VAR2 VAR3); SET TESTS;
RETAIN LIKELIHOOD SCORE WALD;
VAR1="&IVAR1";
VAR2="&IVAR2";
VAR3="&IVAR3";

IF _N_=1 THEN LIKELIHOOD=INPUT(CHISQ,12.3);
ELSE IF _N_=2 THEN SCORE=INPUT(CHISQ,12.3);
ELSE IF _N_=3 THEN DO;
WALD=INPUT(CHISQ,12.3);
OUTPUT;
END;
RUN;

%MEND;
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Invoking the Macro

IHS Automotive
DRIVEN BY POLK

```
%LET FMT=C:\AAAWORK\MISUG\VFORMAT.TXT; /* text file for formats */
%LET OUTLIST=C:\AAAWORK\MISUG\LIST.TXT; /* file to capture output window */
%LET OUTLOG=C:\AAAWORK\MISUG\LOG.TXT; /* file to capture log window */
%LET RESULTS=C:\AAAWORK\MISUG\RESULTS.HTML; /* HTML file holds results */

/* MACRO PARAMETERS */
 * VARS = SAS DATASET CONTAINING INDEPENDENT VARIABLE LIST;
 * STUDY = SAS DATASET CONTAINING MODEL-BUILDING DATA;
 * DV = NAME OF DEPENDENT VARIABLE;
%RLOGIT(VARS,MISUG.STUDY,BUYCNT/HHDCNT)
```

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.

Questions/Comments

IHS Automotive
DRIVEN BY POLK

Barbara Keys

Barbara_Keys@polk.com

© 2013, IHS Inc. No portion of this presentation may be reproduced, reused, or otherwise distributed in any form without prior written consent.