A SAS Macro for Univariate Logistic Regression

Masud Rana

Clinical Research Support Unit, College of Medicine
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5, Canada

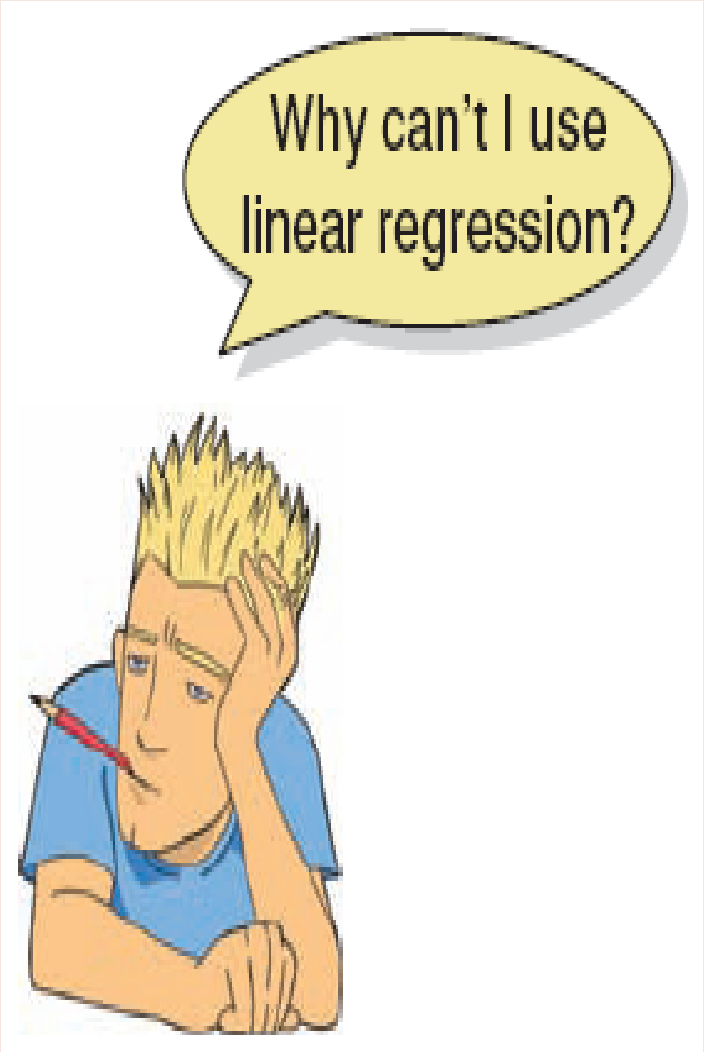
Saskatoon SAS User Group (SUCCESS)

October 24, 2013

Outline

1. Logistic Regression
2. Macro
3. Example

When & Why

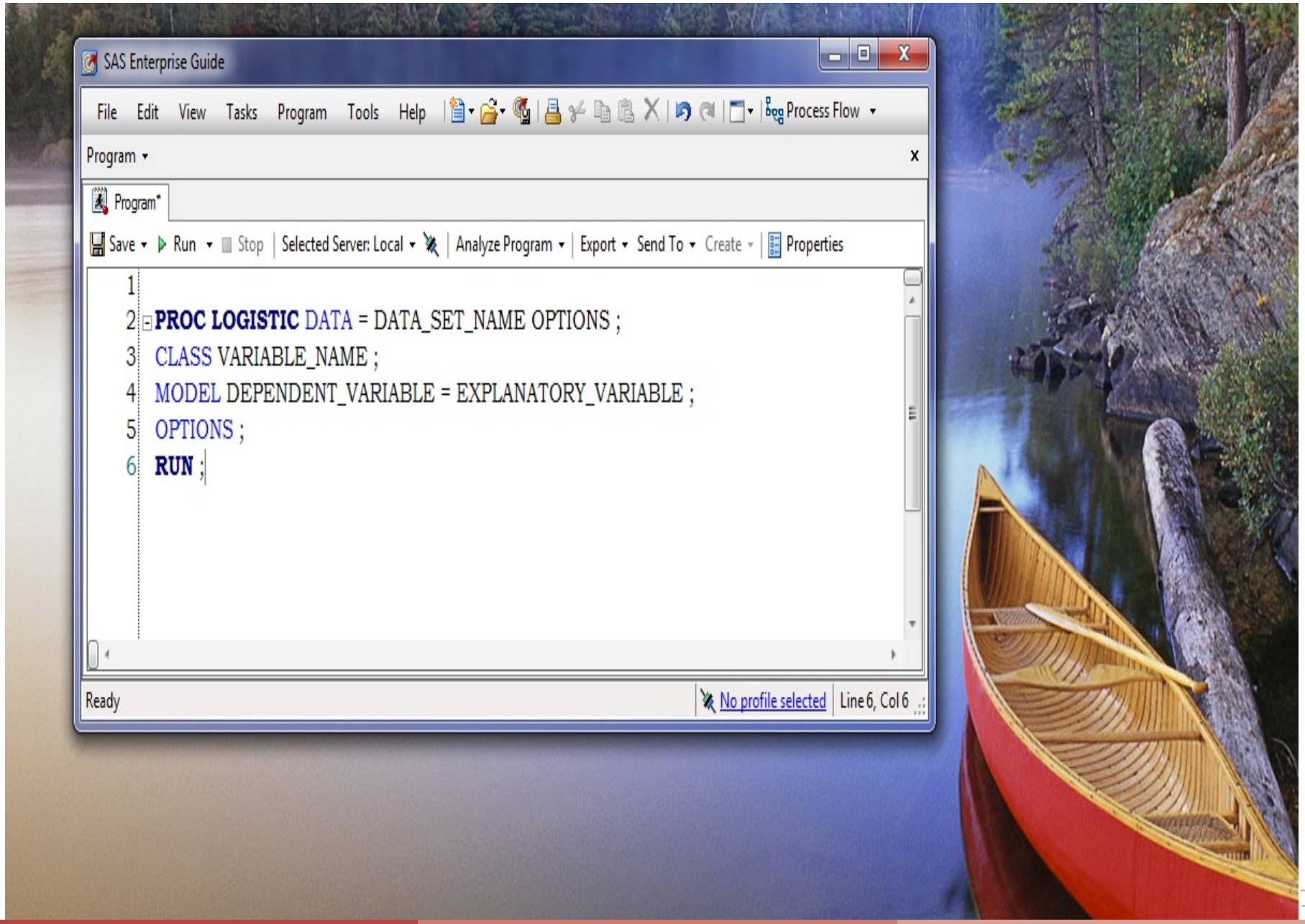


- Response variable is categorical, often binary:
  Yes/No
  Acceptable/Not acceptable
  Phenomena takes place/does not take place
- Violates the assumption of linearity in normal regression.
- Error terms are heteroscedastic and not normal.

Area of Application

- *Health Sciences:* questions about disease: yes or no?
- *Social Sciences:* deals with great deal of dichotomous variables: employed vs unemployed, married vs unmarried, etc.
- Used in economics and marketing to study consumer choice.
- Banks use it to assess credit rating of customers
- Public Opinion Polls
- Used in demand forecasting

PROC LOGISTIC



What is SAS Macro

A macro is a larger piece of a program that can contain complex logic including complete DATA and PROC steps, and macro statements such as *%IF-%THEN -%ELSE* and *%DO-%END*. Macros often - but not always - contain macro variables.[1]

Why use Macro

1. With macros you can make one small change in your program and have SAS echo that change throughout your program.[1]
2. Macros can allow you to write a piece of code and use it over and over again in the same or different program.[1]
3. You can make your program data driven, letting SAS decide what to do based on actual data.[1]

Data

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
9	Cylinders	Num	8		
5	DriveTrain	Char	5		
8	EngineSize	Num	8		Engine Size (L)
10	Horsepower	Num	8		
7	Invoice	Num	8	DOLLAR8.	
15	Length	Num	8		Length (IN)
11	MPG_City	Num	8		MPG (City)
12	MPG_Highway	Num	8		MPG (Highway)
6	MSRP	Num	8	DOLLAR8.	
1	Make	Char	13		
2	Model	Char	40		
4	Origin	Char	6		
3	Type	Char	8		
13	Weight	Num	8		Weight (LBS)
14	Wheelbase	Num	8		Wheelbase (IN)

Data (Contd.)

SAS Enterprise Guide

File Edit View Tasks Program Tools Help Process Flow

Code

Program* Log

Save Run Stop Selected Server: Local (Connected) Analyze Program Export Send To Create

```
DATA CARS ;
SET SASHELP.CARS ;
IF (MSRP LE 32700) THEN Price = 'Cheap';
ELSE Price= 'Costly' ;
IF (MPG_City  LE 20) THEN Efficiency = 'No';
ELSE Efficiency= 'Yes' ;
IF (EngineSize  LE 3) THEN Size = 'Small';
ELSE Size= 'Big' ;
IF (Type  EQ 'Sedan') THEN Sedan = 'Yes' ;
ELSE Sedan= 'No' ;
IF (Origin  EQ 'USA') THEN USA = 'Yes';
ELSE USA= 'No' ;
IF (Origin  EQ 'Asia') THEN Asia = 'Yes';
ELSE Asia= 'No' ;
IF (Origin  EQ 'Europe') THEN Europe = 'Yes';
ELSE Europe= 'No' ;
IF (DriveTrain  EQ 'Front') THEN FrontWheel = 'Yes';
ELSE FrontWheel='No' ;
DROP MSRP MPG_City EngineSize Make Model Type DriveTrain Origin ;
RUN ;
```

Ready No profile selected Line 37, Col 20

What predicts what?

- *Costly or Not*
- *Efficient or Not*
- *Big or Small*
- *Front Wheel Drive or Not*

- *Response Variable:* 4
- *Explanatory Variable:* 11
- *Total Model:* 44

Macro Code

```
/*
This macro runs univariate logistic regression for any number of
outcomes and predictors.It summarizes parameter estimates from
different models into two data files: one for numeric and one
for character variables. Arguments are:
DS=Input Dataset
Outcome=Output Variable(s)
Numvar=One numeric variable name to generate summary data file
CHARVAR=One Character Variable to generate summary data file
Outvar=One output variable name if there are more than one Outcomes
*/

%MACRO LOGIS(DS=, OUTCOME=,NUMVAR=,
                    CHARVAR=, OUTVAR=) ;
DATA _null_;
%* Count number of variables in the ConvertVars parameter ;
CALL SYMPUTX("novars", count("&OUTCOME", "") + 1, "G");
RUN;
%PUT Number of Outcome Variables Count is &novars;
DATA _DSin (DROP=&OUTCOME) ;
SET &DS ;
RUN ;
%INCLUDE 'Z:\SAS Work\Macro\Variable_List.sas' ;
%LST(_DSin) ;
%PUT X=&X ;
%PUT Y=&Y ;
%LET Outvars=1 ;
%DO %WHILE(&Outvars LT (&novars + 1));
%LET newvar = %SCAN(&OUTCOME, &Outvars);
%PUT Processing variable is &newvar;
ODS OUTPUT ParameterEstimates = _SUMMARYC ;
PROC LOGISTIC DATA=&DS  DESCENDING ;
CLASS &CHARVAR ;
MODEL &OUTVAR = &CHARVAR /
RSQUARE LACKFIT ;
RUN;
QUIT;
ODS OUTPUT CLOSE ;
DATA _SUMMARYC ;
LENGTH VARIABLE $ 32 ;
SET _SUMMARYC ;
DELETE  ;
RUN ;
ODS OUTPUT ParameterEstimates = _SUMMARYN ;
```

Macro Code (Contd.)

```
PROC LOGISTIC DATA=&DS  DESCENDING ;
MODEL &OUTVAR = &NUMVAR /
RSQUARE LACKFIT ;
RUN;
QUIT;
ODS OUTPUT CLOSE ;
DATA _SUMMARYN ;
LENGTH VARIABLE $ 32 ;
SET _SUMMARYN ;
DELETE ;
RUN ;
%LET DSID=%SYSFUNC(OPEN(_DSin, i));
%LET K=1 ;
%LET VAR = %SCAN(&X, &K) ;
%LET L=1 ;
%LET VART = %SCAN(&Y, &L) ;
%DO %WHILE ("&VAR" NE "") ;
%IF ("&VART" EQ "C")
%THEN %DO ;
ODS OUTPUT ParameterEstimates = _peC&Outvars&K
                OddsRatios = _orC&Outvars&K ;
PROC LOGISTIC DATA=&DS  DESCENDING ;
CLASS &VAR ;
MODEL &newvar = &VAR /
RSQUARE LACKFIT ;
TITLE1 "Logistic Regression for Predicting &newvar with &VAR (Character) " ;
RUN;
QUIT;
ODS OUTPUT CLOSE ;
DATA _peC&Outvars&K ;
LENGTH VARIABLE $ 32 ;
SET _peC&Outvars&K ;
RUN ;
PROC APPEND BASE = _SUMMARYC DATA = _peC&Outvars&K FORCE ;
RUN ;
%END ;
%ELSE %IF ("&VART" EQ "N")
%THEN %DO ;
ODS OUTPUT ParameterEstimates = _peN&Outvars&K
                OddsRatios = _orN&Outvars&K ;
PROC LOGISTIC DATA=&DS  DESCENDING ;
MODEL &newvar = &VAR /
RSQUARE LACKFIT ;
TITLE1 "Logistic Regression for Predicting &newvar with &var (Numeric) " ;
```

Macro Code (Contd.)

```
RUN;
QUIT;
ODS OUTPUT CLOSE ;
DATA _peN&Outvars&K ;
LENGTH VARIABLE $ 32 ;
SET _peN&Outvars&K ;
RUN ;
PROC APPEND BASE = _SUMMARYN DATA = _peN&Outvars&K FORCE ;
RUN ;
%END ;
%LET K = %EVAL(&K + 1) ;
%LET VAR = %SCAN(&X, &K) ;
%LET L = %EVAL(&L + 1) ;
%LET VART = %SCAN(&Y, &L) ;
%END ;
TITLE1 "Summary of Parameter Estimates for &newvar for Numeric
Variables" ;
PROC PRINT DATA=_SUMMARYN ;
RUN ;
TITLE1 "Summary of Parameter Estimates for &newvar for Character
Variables" ;
PROC PRINT DATA=_SUMMARYC ;
RUN ;
PROC DATASETS LIB=WORK NOLIST ;
DELETE _SUMMARYC _SUMMARYN ;
QUIT ;
RUN ;
%LET Outvars = %EVAL(&Outvars +1);
%PUT Updated vars loop count &Outvars;
%END;
%LET RC=%SYSFUNC(CLOSE(&DSID)) ;
%MEND LOGIS ;
```

Macro Code (Contd.)

SAS Enterprise Guide

File Edit View Tasks Program Tools Help

Code

Program* Log

Save Run Stop | Selected Server: Local (Connected) | Analyze Program

```
%LOGIS(
        DS=CARS,
        OUTCOME= Price Efficiency Size FrontWheel ,
        NUMVAR=Cylinders,
        CHARVAR=Sedan,
        OUTVAR=Price
        );
```

Ready | No profile selected | Line 25, Col 23

References

Susan J. Slaughter and Lora D. Delwiche
SAS Macro Programming for Beginners.
SUGI 29.


THANK YOU