

SUMMARIZATION OF LOGISTIC REGRESSION RESULTS WHEN THERE ARE MANY DEPENDENT VARIABLES AND A COMMON SHARED SET OF EXPLANATORY VARIABLES.

John T. Jones, M.A., M.S.
Department of Biostatistics and Data Management
PharmaResearch Corporation, Wilmington, NC

ABSTRACT

Side by side comparisons of an univariate and multivariate estimate reveal the significance and robustness of that estimate. Using the SAS® logistic procedure and the stepwise regression technique, macro language, data step logic, and the report procedure, a program can be written that succinctly summarizes the univariate and multivariate parameter estimates, permitting side by side comparison, for many dependent variables that share a common set of explanatory variables, in a clear easily interpretable table. This technique can be extended to other regression methods.

INTRODUCTION

A professor of mine once defined a statistic as the summary of some volume of data without the loss of information. What I missed in one of his lectures was how to summarize some volume of statistics without the loss of information.

Business today processes huge amounts of data. Making sense not just of a statistic but of all the statistics that are generated is a difficult task often if not always left to the statisticians and SAS programmers.

This paper presents a methodology of succinctly and autonomously summarizing a large number of logistic regression results, for a given alpha level, while maintaining a level of statistical integrity. The example will use 120 univariate regressions and 8 stepwise multivariate regressions for a total of 128 logistic regressions arranged entirely on a single sheet of paper.

The follow text comprises the programming methodology, table examples, and a SAS program example.

METHODOLOGY

1) Construct the analysis dataset.

Prepare a dataset that contains all of the dependent and independent variables. Any transformations or created class variables would be finished and part of this dataset. This dataset is global and comprehensive containing all the variables of interest.

2) Macro uni(depen=, indep=, alpha=)

Design a macro that loops each dependent variable over every independent variable. This fragment would fit the logistic regression, output the list to a dump file, read this dump file into a SAS dataset, keep the relevant statistics, add pertinent `_type_` by variables, and accumulate the results of each dependent variable.

3) Macro mul(depen=, indep=, alpha=)

Design a macro that loops each dependent variable over the entire set of independent variables. This fragment would fit the logistic regression, output the list to a dump file, read this dump file into a SAS dataset, keep the relevant statistics, add pertinent `_type_` by variables, and accumulate the results of each dependent variable.

4) Combine the two datasets

Take the datasets from the uni and mul macros which accumulated the logistic results and combine them appropriately.

5) Produce the table

Employ the report procedure to generate the table.

TABLE 00.0
SUMMARY OF LOGISTIC REGRESSIONS[1]
PARAMETER ESTIMATES SIGNIFICANT AT P-VALUE $\leq$.05

Independent Var.	DEPENDENT1		DEPENDENT2		DEPENDENT3		DEPENDENT4		DEPENDENT5		DEPENDENT6		DEPENDENT7		DEPENDENT8	
	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise
Independent1																
Independent2																
Independent3																
Independent4																
Independent5																
Independent6																
Independent7																
Independent8																
Independent9																
Independent10																
Independent11																
Independent12																
Independent13																
Independent14																
Independent15																
Independent16																

[1] Tabulation of odds ratios

Source: PharmaResearch Corporation, 00:00:00 on WED, 01JAN97 :: C:\My Programs\SESUG97\logtab_blk.sas...(JTJones)

Bringing Science to Development

PharmaResearch Corporation

Project: 00-000-00
Population: All Subjects

TABLE 00.1
SUMMARY OF PHARMACOECONOMIC LOGISTIC REGRESSIONS[1]
PARAMETER ESTIMATES SIGNIFICANT AT P-VALUE <= 0.1

Predictors	GLOBAL EVALUATION		CONFIRMED DIAGNOSIS		CHANGE IN BODY SIGN SYMPTOM		CHANGE IN SEVERITY INDEX		CHANGE IN SUBJECT PERCEIVED SEVERITY		TREATMENT		TREATMENT WORTH-WHILE		TOTAL COSTS	
	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise
Treatment Group			3.8	3.9					2.1		n/a	n/a	n/a	n/a	12.8	14.3
Prior Treatment Use			4.3	11.8	1.9	2.5					0.5	0.1				
Duration of ACD	0.4	0.5	0.7		0.5	0.5			0.5	0.4	0.6	0.6				2.0
Intensity of Treatment Use			2.1								6.0	4.7	1.9	2.0	4.1	2.7
Gender							0.5									1.9
Ethnic Origin									0.5	0.4						
Age													1.7	1.9	1.7	1.7
Severe Subjects	0.6	0.6					3.1	3.4	7.7	7.5	1.7		1.8	1.9	1.7	
Other Physicians Seen											3.4	2.5			6.3	
First Visit to Investigator	2.4	2.0	2.0	1.9	1.7	1.9	2.1						3.2	3.3		
No. Past Visits for ACD				0.6			1.9				2.3				3.5	
No. Past Physicians Seen			1.8		1.6		3.9	3.8	1.8		3.1		2.8		4.1	2.9
Biopsy											3.6				4.4	3.6
PUVA, UVA, UVB Therapy	0.3	0.5	0.2	0.1			0.2	0.1			0.5	0.1			2.9	11.6
Total Costs			2.6	2.1					2.3	2.0	12.8	15.9	2.2		n/a	n/a

[1] Tabulation of odds ratios

Source: PharmaResearch Corporation., 9:08:56 on THU, 21NOV96 :: C:\My Programs\SESUG97\logtab_exml.sas... (JITJones)

TABLE 00.2
SUMMARY OF PHARMACOECONOMIC LOGISTIC REGRESSIONS[1]
PARAMETER ESTIMATES SIGNIFICANT AT P-VALUE <= 0.2

Predictors	GLOBAL EVALUATION		CONFIRMED DIAGNOSIS		CHANGE IN BODY SIGN SYMPTOM		CHANGE IN SEVERITY INDEX		CHANGE IN SUBJECT PERCEIVED SEVERITY		TREATMENT		TREATMENT WORTH-WHILE		TOTAL COSTS	
	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise	Uni- vari	Step wise
Treatment Group			3.8	4.0	0.7				2.1		n/a	n/a	n/a	n/a	12.8	14.3
Prior Treatment Use			4.3	11.6	1.9	2.5					0.5	0.1				1.8
Duration of ACD	0.4	0.5	0.7		0.5	0.5			0.5	0.4	0.6	0.6				2.1
Intensity of Treatment Use			2.1								6.0	4.8	1.9	1.7	4.1	2.6
Gender							0.5					0.6	0.6	0.5	1.4	1.8
Ethnic Origin								0.4	0.5	0.4			2.6	2.9		
Age													1.7			1.7
Severe Subjects	0.6	0.6					3.1	3.4	7.7	7.5	1.7		1.8			1.7
Other Physicians Seen											3.4	1.9			6.3	3.4
First Visit to Investigator	2.4	2.0	2.0	1.9	1.7	1.9	2.1		1.5		1.3		3.2	3.0	1.4	2.1
No. Past Visits for ACD				0.6			1.9	1.8			2.3					3.5
No. Past Physicians Seen	1.5		1.8		1.6		3.9	3.6	1.8		3.1	1.6	2.8		4.1	
Biopsy				0.4							3.6				4.4	4.3
PUVA, UVA, UVB Therapy	0.3	0.5	0.2	0.1	0.4		0.2	0.1			0.5	0.2			2.9	12.6
Total Costs			2.6	2.2					2.3	2.0	12.8	16.9	2.2	2.1	n/a	n/a

[1] Tabulation of odds ratios

Source: PharmaResearch Corporation.,10:08:56 on THU, 21NOV96 :: C:\MyPrograms\SESUG97\logtab_exm2.sas... (JTIJones)

```

*--{ANALYSIS: dataset location}--;
libname pharm 'w:\biostat\... \pharmeco';

*--{LEFT: logistic list output dump file}--;
options nocenter;
filename dump 'w:\biostat\... \dump.txt';

*--{UNIVARIATE: reg. and data processing}--;
%Macro uni(dep=, ind=, alpha=.1);

*--{ROUTE: the list}--;
proc printto file=dump new; run;

*--{UNIVARIATE: regression}--;
proc logistic data=pharm.anal.dat descending;
model &dep=&ind;
run;

*--{RESET:}--;
proc printto; run;

*--{READ: the list}--;
data infile;
length str1 $100;
infile dump length=linelen;
input str1 $varying100. linelen;
str1=trim(left(str1));
if (substr(str1,1,8)='-2 LOG L') then output;
if (substr(str1,1,1)='X') then output;
run;

*--{RELEVANT: output}--;
data cleanup(keep=mod depen indep
               mod_p est stderr est_p odds);
length mod $1 depen indep $8
       mod_p est stderr est_p odds 8;
set infile end=eof;
retain mod ' ' mod_p est stderr est_p odds .;
if (substr(str1,1,8)='-2 LOG L') then mod_p=
    input(compress(scan(str1,10,' '),'(p=)'),8.);
if (substr(str1,1,1)='X') then do;
*--{desired: statistics}--;
est =input(scan(str1,3,' '),8.);
stderr=input(scan(str1,4,' '),8.);
est_p =input(scan(str1,6,' '),8.);
odds =input(scan(str1,8,' '),8.);
*--{unstable:}--;
if (odds in(999.)=1) then delete;
*--{threshold:}--;
if (est_p > &alpha ) then delete;
end;
*--{type: variables}--;
depen=upcase("&dep");
indep=upcase("&ind");

```

```

mod='U';
if (eof=1) then output;
run;

*--{ACCUMULATE:}--;
data cum_uni;
%if (&loop=1) %then %do; set cleanup; %end;
%else %do; set cum_uni cleanup; %end;
run;

*--{SCOUR:}--;
proc datasets; delete infile cleanup; run;

%Mend uni;

*--{DEPENDENT: variables}--;
%let dep1=xglob; %let dep2=xdmdiag;
%let dep3=xchg_bss; %let dep4=xsev_df;
%let dep5=xsev52_c; %let dep6=xpatchgr;
%let dep7=xpatsat; %let dep8=xbrand;

*--{INDEPENDENT: variables}--;
%let ind1=xpatchgr; %let ind2=xpatch;
%let ind3=xdur; %let ind4=xinvno;
%let ind5=xsex; %let ind6=xrace;
%let ind7=xage; %let ind8=xsev;
%let ind9=xanyphys; %let ind10=xfirphys;
%let ind11=xvisno; %let ind12=xphysno;
%let ind13=xgdbiop; %let ind14=xuv_ther;
%let ind15=xbrand;

*--{CALL: univariate regressions}--;
%Macro uni_loop; %let loop=1;
%do dep=1 %to 8;
%do ind=1 %to 15;
%uni(dep=&dep&dep, ind=&ind&ind);
%let loop=%eval(&loop+1);
%end;
%end;
%Mend uni_loop;
%uni_loop;

*--{MULTIVARIATE: reg and data process.}--;
%Macro mul(dep=, ind=, alpha=.1);

*--{ROUTE: the list}--;
proc printto file=dump new; run;

*--{MULTIVARIATE: regression}--;
proc logistic data=pharm.anal_vis descending;
model &dep=&ind/selection=stepwise
      sle=&alpha sls=&alpha;
run;

```

```

*--{RESET:}--;
proc printto; run;

*--{READ: the list}--;
data infile;
length str1 $100 indep $8;
infile dump length=linelen;
input str1 $varying100. linelen;
str1 =trim(left(str1));
indep=substr(str1,1,8);
if (substr(str1,1,8)='-2 LOG L') then output;
if (substr(str1,1,1)='X') then output;
run;

*--{RETAIN: the last model statistics}--;
data last;
set infile;
by indep notsorted;
if (last.indep=1);
run;

*--{RELEVANT: output}--;
data cleanup(keep= mod depen indep mod_p est
                stderr est_p odds);
length mod $1 depen indep $8 mod_p est stderr
        est_p odds 8;
set last end=eof;
by indep notsorted;
retain mod ' ' mod_p .;
if (substr(str1,1,8)='-2 LOG L') then mod_p=
input(compress(scan(str1,10,' '),'(p=)'),8.);
if (substr(str1,1,1)='X') then do;
*--{desired: statistics}--;
est =input(scan(str1,3,' '),8.);
stderr=input(scan(str1,4,' '),8.);
est_p =input(scan(str1,6,' '),8.);
odds =input(scan(str1,8,' '),8.);
array x{5} mod_p est stderr est_p odds;
do i=1 to 5;
if (x{i}=.) then x{i}=U;
end;
end;
*--{ _type_ : variables}--;
depen=upcase("&dep");
mod='S';
if (indep='-2 LOG L') then delete;
if (last.indep=1) then output;
run;

*--{ACCUMULATE:}--;
data cum_mul;
%if (&loop=1) %then %do; set cleanup; %end;
%else %do; set cum_mul cleanup; %end;
run;

```

```

*--{SCOUR:}--;
proc datasets; delete infile last cleanup; run;

%Mend mul;

*--{DEPENDENT: variables}--;
%let dep1=xglob; %let dep2=xdmdiag;
%let dep3=xchg_bss; %let dep4=xsev_df;
%let dep5=xsev52_c; %let dep6=xpatchgr;
%let dep7=xpatsat; %let dep8=xbrand;

*--{INDEPENDENT: variables}--;
%let ind1=xpatchgr xpatch xdur xinvno xsex
        xrace xage xsev xanyphys
        xfirphys xvisno xphysno xdgbiop
        xuv_ther xbrand;

%let ind6=*****/ xpatch xdur xinvno xsex
        xrace xage xsev xanyphys
        xfirphys xvisno xphysno xdgbiop
        xuv_ther xbrand;

%let ind8=xpatchgr xpatch xdur xinvno xsex
        xrace xage xsev xanyphys
        xfirphys xvisno xphysno xdgbiop
        xuv_ther /*****/;

*--{CALL: multivariate regressions}--;
%Macro mul_loop; %let loop=1;
%do dep=1 %to 8;
%if (&dep<6) %then %let ind=&ind1;
%else %if (&dep<8) %then %let ind=&ind6;
%else %let ind=&ind8;
%mul(dep=&dep&dep, ind=&ind);
%let loop=%eval(&loop+1);
%end;
%Mend mul_loop;
%mul_loop;

*--{COMBINE: uni and multivariate results}--;
data table;
set cum_uni cum_mul;
run;

*--{DESCRIBE:}--;
proc format;
value $mod 'S'='Step@wise' 'U'='Uni-@vari';
value $dep
'XGLOB' ='GLOBAL@EVALUATION'
'XDMDIAG' ='CONFIRMED@DIAGNOSIS'
'XCHG_BSS'='CHANGE IN@
                BODY@SIGN@SYMPTOM'
'XSEV_DF'='CHANGE IN@

```

```

SEVERITY@INDEX'
'XPATCHGR'='PATCH@TESTED'
'XSEV52_C'='CHANGE IN@
SUBJECT@PERCEIVED@
SEVERITY'
'XPATSAT'='PATCH@TEST@
WORTH@WHILE'
'XBRAND' ='TOTAL COSTS';
value $ind 'XPATCHGR'='Patch Group'
'XPATCH' ='Prior Patch Use'
'XDUR' ='Duration of ACD'
'XINVNO' ='Intensity of Patch Use'
'XSEX' ='Gender'
'XRACE' ='Ethnic Origin'
'XAGE' ='Age'
'XSEV' ='Severe Subjects'
'XANYPHYS'='Other Physicians Seen'
'XFIRPHYS' ='First Visit to Investigator'
'XVISNO' ='No. Past Visits for ACD'

```

```

'XPHYSNO' ='No. Past Physicians Seen'
'XDGBIOP' ='Biopsy'
'XUV_THER'='PUVA, UVA, UVB
Therapy'
'XBRAND' ='Total Costs';

```

run;

```

*--{LINE:}--;
data _null_;
call symput('undl',repeat('_',128));
run;

```

```

*--{WHERE: when, what, who stamp}--;
%stamp(whoami=JTJones);

```

```

*--{OUTPUT: specific options}--;
options nodate nonumber ls=132 ps=53 nocenter
missing='';

```

```

*--{OUTPUT:}--;
proc report data=table headline missing nowd split='@';
column ("&undl" indep depen,( mod, (odds)));
define indep /group order=data width=30 format=$ind. 'Predictors';
define depen /across order=data width= 8 format=$dep. '';
define mod /across order=data width= 8 format=$mod. '';
define odds /analysis min width= 4 format=4.1 '';
break after indep / skip;
title1 'Project: 00-000-00';
title2 'Population: All Subjects';
title3 'TABLE 00.0';
title4 'SUMMARY OF PHARMACOECONOMIC LOGISTIC REGRESSIONS[1]';
title5 'PARAMETER ESTIMATES SIGNIFICANT AT P-VALUE <= 0.1';
footnote1 '[1] Tabulation of odds ratios';
footnote2;
footnote3 "&stamp";
run;

```

REFERENCES:

Hosmer, D.W., Jr. And Lemeshow, S. (1989). *Applied Logistic Regression*, New York: John Wiley & Sons, Inc.

SAS Institute Inc., *SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2* (1989)

SAS Institute Inc., *SAS® Language: Reference, Version 6, First Edition* (1990)

SAS Institute Inc., *SAS® Macro Language: Reference, First Edition* (1997)

SAS Institute Inc., SAS® Technical Report P-258, *Using the REPORT Procedure in a Nonwindowing Environment*, Release 6.07 (1993)

John T. Jones MS, MA
PharmaResearch Corporation
1410 Commonwealth Drive, Suite 215
Wilmington, NC 28403
(910) 256-2955
JonesJ@PharmaResearch.com