

/\* Univariate risk factor of outcome with PROC FREQ  
DECEMBER 2007

Programmers: Adrien FRANCAIS and Aurélien VESIN  
Statistical Enginners

Albert Bonniot Institute  
Team 11-Outcome of cancer and critical illnesses  
UJF-INSERM U823  
38706 LA TRONCHE CEDEX  
FRANCE  
04 76 54 94 00

The main Macro: %TEST  
Macro to compute frequencies and percentages by outcome variable, included also the Khi2-  
Mantel-Haenszel test  
and present correctly results.

#### MACRO VARIABLES

TABLE : work table to select variables in the model.  
VARINTERET : Outcome variable coded as 0 or 1.  
LISTE : variable to test if it is a risk factor of outcome variable (varinteret).  
It must be binary variables(0 or 1) or variables in classes.  
SORTIE : table which contains results. This table is printed in the Results section  
FORMAT : put the name of format for the appearance of results  
For example, I put 'codes.' for my data

1-MACRO included in the main program : %CREATIONVARBINAIRES  
Macro to create several binary variables for variables in classes

2-MACRO included in the main program: %EFFECTIF  
Macro to compute frequencies and percentages by outcome variable.

Before using this macro, you have to :

- Prepare your database with correct variables
- Compile the macro.
- Execute the macro :

Example to compute on your computer to directly and easily apply the macro:

```
data try;
input dc14jours      sexe transfert dnr sympt @20;
cards;
0      1      0      0      601
```

```

0    1    1    0    602
1    1    1    0    602
1    1    1    0    602
0    1    1    0    605
0    0    1    0    606
0    1    0    0    607
1    0    1    1    608
1    0    1    0    624
1    1    1    0    610
0    0    0    0    602
0    1    1    0    601
0    1    1    0    609
0    1    0    0    609
1    0    1    1    610
0    1    0    0    611
0    0    1    0    612
1    1    0    1    613
0    1    0    1    614
1    1    1    1    615
1    0    1    0    611
1    0    1    0    .
0    0    0    0    .
0    1    1    0    .
0    0    0    0    614
0    0    .    0    613
1    0    1    1    612
1    1    1    1    601
1    0    1    0    601
1    0    1    0    602
0    1    0    0    608
0    1    0    0    608
0    0    0    0    .
1    0    1    0    .
1    0    0    1    .
0    1    0    0    .
0    1    1    1    609
1    1    1    1    .
0    1    1    0    617
0    0    1    0    605

```

```

;
run;

```

```

%test(table=try,varinteret=dc14jours,liste=sexe transfert dnr
SYMPT,sortie=tablesortie,format=);

```

You can see LOG screen for information of algorithm.

In the OUTPUT screen, you have the summary table with risk factors at the final step.

If you use this macro for work to be published please use the following citation:

Francais Adrien and Vesin Aurélien (2007).  
Univariate risk factor of outcome with PROC FREQ.  
\*/

```
/*CODE*/
%macro test(table=,varinteret=,liste=,sortie=,format=);
%let error=0;
%let errormodal=0;
/*some errors messages*/
%if &table = %then %do;
    %put "ERROR: You didn't give the work table : Please enter the name of work
table";%let error=1;
%end;
%if &varinteret = %then %do;
    %put "ERROR: You didn't give outcome variable : Please enter a binary
variable";%let error=1;
%end;
%if &liste = %then %do;
    %put "ERROR: You didn't give potentially variables to compare : Please enter
qualitative variables";%let error=1;
%end;
%if &sortie = %then %do;
    %put "ERROR: You didn't give the name of final table : Please enter the name of the
final table";%let error=1;
%end;

/*to test if the outcome variable is binary*/
%if &table ne and &error=0 %then %do;
%put "We test if the outcome variable is binary";
ods output onewayfreqs=aaa(keep=&varinteret);
title "Test of variable &varinteret";
proc freq data=&table;tables &varinteret;run;
data aaaa;set aaa;id=1;run;
data aaaaa;set aaaa;by id;
retain errormodal 0;
if first.id and &varinteret ne 0 then errormodal=1;
if last.id and &varinteret ne 1 then errormodal=1;
call symput ('errormodal',errormodal);
run;
%end;
%if &errormodal = 1 %then %do;
    %put "ERROR: You didn't give a correct outcome variable : Please enter a binary
variable";%let error=1;
%end;
```

```

%if &error = 1 %then %do;
    %put "PLEASE CORRECT ERROR(S) DETAILED ABOVE";
%end;

/*if there is no error then we start the macro*/
%if &error ne 1 %then %do;

    /*we create binary variables for variables with multiples values*/
    data tableintermediaire;set &table;run;
    data libelle;format variablebis $50. itembis 5. ;run;
    %local varlist j var;
    %global &liste;
    %let varlist = &liste;
    %let j=0; %let listevarbi=;

    %do %while(%scan(&varlist,&j+1,%str( )) ne %str( ));
        %let j = %eval(&j+1);
        %let var = %scan(&varlist,&j,%str( ));
        %creationvarbinaires(variable=&var,varcible=&varinteret);
        /*we calculate the number of possible values of the variable*/
        ods output onewayfreqs=one;
        proc freq data=&table;tables &var;run;
        data sfsdf;set one;keep &var total;
        retain total 0;total=total+1;
        if cumpercent=100 then output;run;
        data fsfd;set sfsdf;call symput ('maxim',total);run;

        /*we create table with correct label of variables*/
        data one2;set one;variablebis=substr(table,7,30);rename
&var=itembis;keep variablebis &var;run;
        data libelle;set libelle one2;run;

        /*we create the list of binary variables to test*/
        data tableintermediaire;set tableintermediaire;
        %let inter=;%let k=0;%let varinterm=%substr(&var,1,8);
        %if &maxim>2 %then %do;
            %do %while (&k<&maxim);
                %let k = %eval(&k+1);
                %let inter=&varinterm&k;
                %let listevarbi=&listevarbi &inter;
            %end;
        %end;
        /*if it's a binary variable, we keep once value : 1*/
        %if &maxim=2 %then %do;
            %let inter=&varinterm.2;
            %let listevarbi=&listevarbi &inter;
        %end;
    run;
%end;

```

```

%let listebinaire=&listevrbi;
%put "The list of variables to test is &listebinaire";
/*we created the good list of binary variables*/
data tablewithbinary;set tableintermediaire;run;
/*We create an empty table to collect results*/
data effectif;format variable $50. items $50. n_0 10. p_0 4.1 n_1 10. p_1 4.1 nb_miss
10.; run;
/*We keep from the list variable by variable*/
%local i var;
%let i=0;
%do %while(%scan(&listebinaire,&i+1,%str( )) ne %str( ));
%let i = %eval(&i+1);
%let var = %scan(&listebinaire,&i,%str( ));
/*for each variable, we apply macro to have frequencies and
percentages*/
%put %upcase("At step &i, we analyse the variable &var");
%effectif(tablewithbinary,&var,&varinteret);
%end;

/*It is the part for the presentation of results for publication like :
frequencies (percentages)*/
%put %upcase("We compress results for the appearance");
data effectif2; set effectif;item=items+0;
resume_&varinteret=compress(n_0)!! ('!!compress(p_0)!!');
resume_not&varinteret=compress(n_1)!! ('!!compress(p_1)!!');
if item in (0,.) then delete;
keep resume_&varinteret resume_not&varinteret nb_miss;
run;

/*we keep Khi-2 de Mantel-Haenszel*/
%put %upcase("We keep Khi-2 de Mantel-Haenszel");
proc sort data=tablewithbinary;by &varinteret;run;
ods output chisq=chisyut;
title 'Khi-2 de Mantel-Haenszel';
proc freq data=tablewithbinary ;
tables (&listebinaire)*&varinteret/chisq;
run;

data chisyut;set chisyut;if statistic="Khi-2 de Mantel-Haenszel";keep prob;
rename prob=testchi2;label prob="p value of Khi2-test";run;

/*We merge 2 tables (frequencies and Chi2) */
data inter;
format resume_not&varinteret $50. resume_&varinteret $50. nb_miss 5. testchi2
6.4;run;

data libelle2;set libelle;if itembis in (0,.) then delete;run;

/*we obtain the final table wih good label*/
data &sortie;

```

```

merge libelle2 inter effectif2 chisyut;
label variablebis="variable"
itembis="modality"
resume_not&varinteret="&varinteret=0"
resume_&varinteret="&varinteret=1";
run;

data tablewithbinary;set tableintermediaire;run;
%put %upcase("The final table with binary variables created is 'tablewithbinary' and is
available in the Explorator section");

title 'Univariate risk factor of outcome';
%put %upcase("The final table is printed in the Results section");

/*we print the results*/
%if &format ne %then %do;
    proc print data=&sortie label noobs;
        format itembis &format;
    run;
%end;
%else %do;
    proc print data=&sortie label noobs;run;
%end;
/*We obtain the final table*/

title 'Delete tables';
proc datasets ;
delete inter effectif2 chisyut effectif aaa aaaa aaaaa
f0 f1 f0f1 temp freq miss missing means;
run;

%end;
quit;
%mend ;

/* MACRO FREQUENCIES AND PERCENTAGES*/
%macro effectif(table,varexp,varcible);
%let num=0;
data temp;set &table;keep &varexp &varcible;run;
data temp;set temp;varexp2=compress(put(&varexp,15.));drop &varexp;run;
data temp;set temp;&varexp=varexp2;drop varexp2;run;

* WE KEEP IN MEMORY FREQUENCIES FOR EACH GROUP;
%put "WE KEEP IN MEMORY FREQUENCIES FOR EACH GROUP";
title 'Frequencies by group';
proc freq data=temp noprint;table &varexp*&varcible /out=freq;run;

data freq;set freq;&varexp=compress(&varexp);run;
data freq;set freq;where &varexp not in (','');run;

```

```

data f0 f1;set freq(drop=percent);if &varcible=0 then output f0;if &varcible=1 then
output f1;
run;

* WE KEEP IN MEMORY MISSING VALUES;
%put "WE KEEP IN MEMORY MISSING VALUES";
ods output OneWayFreqs=miss;
title 'Missing values';
proc freq data=temp ;table &varexp /missing;run;
ods output close;

data miss;set miss;nobis=1;call symput('num',nobis);run;
%if &num=1 %then %do;
    data missing;set miss;variable="&varexp";nb_miss=frequency;
    label nb_miss="Number of missing values";where &varexp in (','');
    keep variable nb_miss;run;
%end;

* PERCENTAGES OF EACH VARIABLE FOR GROUP=0;
%put "WE OBTAIN PERCENTAGES of VARIABLE &varexp FOR GROUP=0";
data f0;set f0;variable="&varexp";items=compress(put(&varexp,20.));n_0=count;drop
&varcible &varexp count ;label n_0="&varcible = 0 (n)" ;run;
proc sort;by variable items;run;
proc means data=f0 noprint;var n_0;output out=means sum=sum;run;
data f0;merge f0 means;run;
data f0;set f0;retain s;if sum>0 then s=sum;p_0=round((n_0/s)*100,.1);label
p_0="&varcible = 0 (%)";drop _type_ _freq_ sum s;

* PERCENTAGES OF EACH VARIABLE FOR GROUP=1;
%put "WE OBTAIN PERCENTAGES OF VARIABLE &varexp FOR GROUP=1";
data f1;set f1;variable="&varexp";items=compress(put(&varexp,20.));n_1=count;drop
&varcible &varexp count ;label n_1="&varcible = 1 (n)";run;
proc sort;by variable items;run;
proc means data=f1 noprint;var n_1;output out=means sum=sum;run;
data f1;merge f1 means;run;
data f1;set f1;retain s;if sum>0 then s=sum;p_1=round((n_1/s)*100,.1);label
p_1="&varcible = 1 (%)";drop _type_ _freq_ sum s;

* MERGE VARIABLES;
%put %upcase("We obtain results of variable &varexp");
data f0f1;merge f0 f1 ;by variable items;run;
%if &num=1 %then %do;
    data f0f1;merge missing f0f1;by variable;run;
%end;

* ADD AT EACH TIME RESULTS;
%put %upcase("We add results of variable &varexp in the table");
data effectif;set effectif f0f1;run;

quit;
%mend;

```

```
%macro creationvarbinaires(variable=,varcible=);
proc sort data=tableintermediaire;by descending &variable ;run;
/*length of variables must be inferior to 8!!!*/
data tableintermediaire;set tableintermediaire;rename &variable=%substr(&variable,1,8);run;

%let varreduit=%substr(&variable,1,8);

/*Proc glmmod produces automatically binary variables*/
proc glmmod data=tableintermediaire noprint
    outdesign=zzz (drop=&varreduit.0)
        prefix=&varreduit
        zerobased;

    class &varreduit;
    model &varcible = &varreduit;
run;

/*we add new binary variables*/
data tableintermediaire;merge tableintermediaire zzz;run;

%mend;
```