Macro to Determine the Significance of the Main Effects of Two Categorical Factors for the PHREG Procedure

Robert J. Gallop, University of Pennsylvania, Philadelphia, PA

ABSTRACT

This poster presents a macro to determine the main effects of two categorical factors in a Two-Factor ANOVA/ANCOVA linear model with a predetermined link function, where the PHREG procedure requires dummy variable coding for each categorical factor. One of the major problems in using a procedure with dummy variable coding is the analysis of the main effects of each categorical variable can not be immediately determined. All estimates are compared to the default category of the dummy variable coding. Analysis of the main effects requires use of the TEST statement, for which coding gets more complicated with the increase in levels of each factor. This macro will allow you to enter: the categorical factor names, the size of each, and the other independent variables. The macro will create the dummy variables for each factor and the interaction between the two factors and generate the results of the test that the main effects for each categorical factor are equal to 0.

There is not always an alternative procedure to avoid dummy coding, i.e. PHREG. For procedures with no alternative procedure this macro is appropriate.

INTRODUCTION

The primary objective of this poster is to present you with a method to handle a Two-Factor ANOVA/ANCOVA survival analysis using the PHREG procedure.

SITUATION

A possible situation to be presented with is an examination of hazard functions controlling for two categorical factors, while controlling for other binary or continuous explanatory variables. There is no guarantee that the difference in the hazard function for any two levels of Factor 1 is the same for each level of Factor 2. Therefore the possibility of an interaction between the different levels of the factors exists and must be modeled (Devore, 1987).

PHREG procedure does not have a CLASS statement, and therefore, does not allow for categorical variables. This creates several problems. First, dummy variables must be assigned for each categorical factor, as well as the interaction between each factor. Second, an analysis of the main effects for each factor and the significance of the interaction must be performed.

When performing the analysis of the main effects for each factor and the significance of the interaction the Type III Sum of Squares will be used. In Type III Sum of Squares, the significance of each effect is adjusted for all other effects possible (SAS, 1990). To perform the analysis, the TEST statement in the PHREG procedure will be used. The most common use of the TEST statement is to test hypotheses about sets of dummy variables, that correspond to the factors in the model (Allison, 1995).

HYPOTHESES

The analysis consists of three hypotheses. The first hypothesis is that there is no difference in hazard functions across the levels of the first factor. The second hypothesis is that there is no difference in hazard functions across the levels of two factors. The third hypothesis is that there is no difference in the change of hazard functions for any two levels of one factor for each level of the other factor (Allison, 1995). Hazard function can be thought of as failure rate (Bain, 1992).

DISCUSSION

Our task is two-fold to mirror the two problems discussed. First, we must assign the dummy variables and second, we must build the TEST statement to perform the Type III Sum of Squares analysis for the significance testing of our three hypotheses.

MACRO SYNTAX

Below is the syntax for the macro to perform the above tasks.

%efftest(data=, reg=, depend=, indep=, specs=, cat1=, cat2=, s_cat1=, s_cat2=);

where

data - specifies the data set you are using.
reg - specifies the SAS procedure used. Any procedures that use the TEST statement are applicable (i.e. REG, PHREG, LOGISTIC).
depend - specifies the dependent variable in the model. The %str() macro function is used for the censoring specification needed in the PHREG procedure.
indep - specifies the list of other explanatory variables in the model. The %str() macro function is used for a list of variables.
specs - specifies any options desired for the model statement, as well as statements other than the TEST statement in the designated procedure. Options are preceded by a /.

Statements may include any of the following: OUTPUT, BASELINE, BY, PLOT, PRINT, WEIGHT. The end of the options and statements are separated by the usual semicolon.
cat1 - first categorical factor.
cat2 - second categorical factor.
s_cat1 - size of the first categorical factor.
s_cat2 - size of the second categorical factor.

The SAS code for the macro is attached (Attachment 1).

ARE THE RESULTS PRODUCED CORRECT?

The macro uses the TEST statement to perform the significance testing of the main effects and interaction. In a Two-Factor ANCOVA, the GLM procedure is used to perform the analysis instead of the REG procedure which requires dummy variable coding. The syntax usage of the TEST statement in the REG, LOGISTIC, and PHREG is identical (Allison, 1995). Therefore, by acquiring equal main effect results and equal interaction results in GLM procedure and REG.
procedure (through the macro), proves the macro is acquiring the correct result. Attachment 2 contains sample data set, GLM syntax, GLM output, MACRO syntax, and MACRO output. Attachment 2 illustrates the results are the same.

WARNINGS

First, the levels of the categorical variables must be 0, 1, 2, 3, . . . , N-1, where N is the number of levels for the variable. If the levels do not match the above specification, recoding of the variable is required.

Second, in version 6.11 and version 6.12 of the SAS software, the TEST statement is limited to 100 fields. Due to this, constraints are placed on the size of the categorical factors.

<table>
<thead>
<tr>
<th>Size of the Smaller Factor</th>
<th>Maximum Size of the Larger Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

If you exceed the above restrictions, the macro will cause an invalid page fault in Kernel32.DLL, which will exit you from the SAS system. SAS Technical support explains this problem has been corrected for Version 7 (SAS, 1991-1997).

Third, you must ensure there are no empty cells in the contingency table of your two categorical factors.

EXAMPLE with PHREG

An investigation of failure rates across 5 levels of site and 3 levels of treatment condition while controlling for job status of patients (employed v. unemployed) and age of patients was conducted on a sample of 95 patients. Time was measured as days in the study. Attachment 3 contains the data, macro syntax, and macro output. As Attachment 3 illustrates, there is no difference of failure rate over the 5 sites, no difference of failure rate over the 3 levels of treatment, and no difference in failure rates between any two sites for each treatment condition, while controlling for age and employment status of the patient.

CONCLUSIONS

This poster has shown you how to use the macro to examine the significance of the main effects of each of the two categorical factors in the model, as well as the significance of the interaction between the two categorical factors. This poster also showed this macro is applicable for other procedures such as LOGISTIC, and REG.

REFERENCES


ACKNOWLEDGEMENTS

The author would like to express his appreciation to the following people for their assistance in this paper:

Jesse Chittams, Dept. of Psychiatry, University of Pennsylvania, Philadelphia, PA

Elizabeth McCalmont, Dept. of Psychiatry, University of Pennsylvania, Philadelphia, PA

Edward Routten, SAS Institute Inc., Cary, NC

TRADEMARK

SAS, SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

Robert J. Gallop
Dept. of Psychotherapy
University of Pennsylvania
3600 Market Street, Room 706
Philadelphia, PA 19104
email: gallop@landru.cpr.upenn.edu
Phone: (215) 349-5646
ATTACHMENT 1: MACRO ALGORITHM

/*------------------------------------------*/
/* eff1pt2 macro and eff1pt1 macro form the */
/* syntax for the TEST Statement.           */
/*------------------------------------------*/

%macro eff1pt2(s_cat1=&s_cat1, s_cat2=&s_cat2);
  %do i= 1 %to (&s_cat1 - 2);
    ,
    &s_cat2*c1i +
    %do j= 1 %to (&s_cat2 - 1);
    c1i.c2j +
    %end;
  0 =
  %let r= %eval(i +1); &s_cat2*c1r +
  %do j= 1 %to (&s_cat2 - 1);
  c1r.c2j +
  %end;
  0
  %end;
%end eff1pt2;

%macro eff1pt1(s_cat1=&s_cat1, s_cat2=&s_cat2);
  &s_cat2*c1i +
  %do i= 1 %to (&s_cat2 - 1);
  c1i.c2i. +
  %end;
  0
  %if &s_cat1 ne 2 %then eff1pt2(s_cat1=&s_cat1, s_cat2=&s_cat2);
%end eff1pt1;

/*-----------------------------------------------------*/
/* dummy1 macro forms the dummy variable coding for */
/* each factor and the interaction between the two */
/* factors.                                       */
/*-----------------------------------------------------*/

%MACRO dummy1(DATA=, CAT1=1, CAT2=2, S_CAT1=1, S_CAT2=);
DATA &DATA;
SET &DATA;

%DO I= 1 %TO (&S_CAT1 - 1);
  IF &CAT1=1I THEN C1I. = 1;
  IF &CAT1 NE &I AND &CAT1 GE 0 THEN C1I. = 0;
  IF &CAT1 < 0 THEN C1I. = . ;
  %LET C1C1I.=&I;
%END;

%DO J= 1 %TO (&S_CAT2 - 1);
  IF &CAT2=1J THEN C2J. = 1;
  IF &CAT2 NE &J AND &CAT2 GE 0 THEN C2J. = 0;
  IF &CAT2 < 0 THEN C2J. = . ;
  %LET C2C2J=&&J;
%END;

%DO I= 1 %TO (&S_CAT1 - 1);
  %DO J= 1 %TO (&S_CAT2 - 1);
    C1I.C2J. = C1I.*C2J. ;
  %END;
%END;
%END;
%MEND dummy1;

/* ------------------------------------------------------*/
/* Blgd1 macro inspects the equality of the hazard */
/* functions over the levels of the first factor. */
/* ------------------------------------------------------*/

%MACRO Blgd1(DATA=,REG=,
   DEPEND=, INDEP=, SPECS=,
   S_CAT1=,S_CAT2=);
   %global SPLITINT;
   %LET SPLITINT=
   %DO I= 1 %TO (&S_CAT1 - 1);
   %DO J= 1 %TO (&S_CAT2 - 1);
   %LET SPLITINT= &SPLITINT C1&I.C2&J.
   %LET C1C1C1=&I;
   %LET C2C2C2=&J;
   %END;
   %END;

PROC &REG DATA=&DATA ;
MODEL &DEPEND = C11-C1&C1C1 C21-C2&C2C2 C2 &INDEP &SPLITINT &SPECS ;

CAT1_EF: test 0= %effipt1(s_cat1=&S_CAT1,s_cat2=&S_CAT2) ;
QUIT;

data &data;
set &data;
drop C11-C1&C1C1 C21-C2&C2C2 &SPLITINT ;
r
run;
%MEND Blgd1;

/* ------------------------------------------------------*/
/* Blgd2 macro examines the equality of the hazard */
/* function over the levels of the second factor, as */
/* well as an inspection of the significance of the */
/* interaction. */
/* ------------------------------------------------------*/

%MACRO Blgd2(DATA=,REG=,
   DEPEND=, INDEP=, SPECS=,
   S_CAT1=,S_CAT2=);
   %LET SPLITIN2=
   %DO I= 1 %TO (&S_CAT1 - 1);
   %DO J= 1 %TO (&S_CAT2 - 1);
   %LET SPLITIN2= &SPLITIN2 C1&I.C2&J.
   %LET C1C1C1=&I;
   %LET C2C2C2=&J;
   %END;
   %END;

PROC &REG DATA=&DATA ;
MODEL &DEPEND = C11-C1&C1C1 C21-C2&C2C2 C2 &INDEP &SPLITIN2 &SPECS ;

CAT2_EF: test 0= %effipt1(s_cat1=&S_CAT1,s_cat2=&S_CAT2) ;
INTER: TEST
   %DO I= 1 %TO (&S_CAT1 - 1);
   %DO J= 1 %TO (&S_CAT2 - 1);

4
C1&I.C2&J.
%IF &I NE %EVAL(&S_CAT1 - 1) OR &J NE %EVAL(&S_CAT2 - 1) %THEN %STR(1);
%END;

%END;
;
QUIT;

data &data;
set &data;
drop C11-C1&C1C1 C21-C2&C2C2 &SPLITIN2 ;
run;

%MEND Bldg2;

_MACRO EFFTEST macro groups each 4 separate _/
_MACRO macros into one full _ macro. */
_MACRO _

%macro efftest(DATA=,REG=,DEPEND=, INDEP=,SPECS=, CAT1=, CAT2=, S_CAT1=, S_CAT2=);
options formdlim='';
%dummy1(DATA=&data, CAT1=&cat1, CAT2=&cat2, S_CAT1=&_cat1, S_CAT2=&_cat2);
%Bldg1(DATA=&data,REG=&reg,DEPEND=&depend, INDEP=&indep,SPECS=&specs, S_CAT1=&_cat1, S_CAT2=&_cat2);
%dummy1(DATA=&data, CAT1=&cat1, CAT2=&cat1, S_CAT1=&_cat2, S_CAT2=&_cat1);
%Bldg2(DATA=&data,REG=&reg,DEPEND=&depend, INDEP=&indep,SPECS=&specs, S_CAT1=&_cat2, S_CAT2=&_cat1);
%mend efftest;
Example

This example appears in GLM Procedure Example 3: Unbalanced ANOVA for Two-WAY Design with Interaction (SAS, 1990).

This example uses data from Kutner (1974) to illustrate a two-way analysis of variance. The original data source is Afifi and Azen (1972).

Data Set

data a;
input drug disease @;
do j=1 to 6;
    input y @;
    output;
endo;
cards;
1 1 42 44 36 13 19 22
1 2 33 . 26 . 33 21
1 3 31 -3 . 25 25 24
2 1 28 . 23 34 42 13
2 2 34 33 31 . 36
2 3 26 28 32 4 16
3 1 19 12 12 -5 16 15
3 2 . 11 9 7 1 -6
3 3 21 1 9 3 .
4 1 24 . 9 22 -2 15
4 2 34 33 31 . 36
4 3 22 7 25 5 12 .
;run;

ANOVA Syntax

proc glm;
class drug disease;
model y=drug disease drug*disease / ss3;
quit;

GLM Output

Dependent Variable: Y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRUG</td>
<td>3</td>
<td>2977.47</td>
<td>999.15</td>
<td>9.05</td>
<td>0.0001</td>
</tr>
<tr>
<td>DISEASE</td>
<td>2</td>
<td>415.87</td>
<td>207.94</td>
<td>1.88</td>
<td>0.1637</td>
</tr>
<tr>
<td>DRUG*DISEASE</td>
<td>6</td>
<td>707.26</td>
<td>117.88</td>
<td>1.07</td>
<td>0.3958</td>
</tr>
</tbody>
</table>

Macro Syntax

data a;
set a;
drug = drug - 1; ***Recode levels to 0,1,2,3***;
disease = disease - 1; ***Recode levels to 0,1,2***;
run;
%efftest(data=a, reg=reg, depend=y,
specs = %str(/noprint;), cat1=drug, cat2=disease,
s_cat1=4, s_cat2=3);

Macro Output

Dependent Variable: Y

Test: CAT1_EF Numerator: 999.1573 DF: 3 F value: 9.0460
                  Denominator: 110.4525 DF: 46 Prob>F: 0.0001

Test: CAT2_EF Numerator: 207.9965 DF: 2 F value: 1.8826
                  Denominator: 110.4525 DF: 46 Prob>F: 0.1637

Test: INTER Numerator: 117.8777 DF: 6 F value: 1.0672
                  Denominator: 110.4525 DF: 46 Prob>F: 0.3958
As is illustrated, both methods acquire the same results. There is a significant difference among the four drugs. The DISEASE effect and DRUG*DISEASE interaction are not significant. Therefore, the macro does create the necessary dummy variables and syntax for the TEST statement to inspect the main effects of two categorical factors and the interaction between the two factors, while controlling for other variables in the model. Since the TEST statement syntax is the same in the REG procedure and PHREG procedure, thus, the macro will properly assess our three hypotheses for the survival model (Allison, 1995).
ATTACHMENT 3 – EXAMPLE OF MACRO FOR PHREG PROCEDURE

Sample Data is a random sample of 95 observations provided from Pilot System File Manual (NIDA Collaborative Cocaine Treatment Study, 1996).

Sample Data

<table>
<thead>
<tr>
<th>Time</th>
<th>Censor</th>
<th>Job</th>
<th>Age</th>
<th>Site</th>
<th>TX</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>41</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>29</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>61</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>114</td>
<td>0</td>
<td>1</td>
<td>35</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>0</td>
<td>36</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>35</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>Texas</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>136</td>
<td>0</td>
<td>1</td>
<td>41</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>40</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>26</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>33</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>61</td>
<td>0</td>
<td>1</td>
<td>38</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>0</td>
<td>39</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>67</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>140</td>
<td>0</td>
<td>1</td>
<td>28</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>40</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>1</td>
<td>36</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>0</td>
<td>1</td>
<td>27</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>76</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>52</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>1</td>
<td>31</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>39</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>63</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>90</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>35</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>38</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>62</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>29</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>36</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>26</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>44</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>76</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Macro Syntax

%efftest(data=test1, reg=%str(phreg), depend=%str(time*censor(1)), indep=%str(job age),
specs=%str(/ties=exact), cat1=site, cat2=txcond, s_cat1=5, s_cat2=3);

Macro Output

Linear Hypotheses Testing

<table>
<thead>
<tr>
<th>Label</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT1_EF</td>
<td>6.5286</td>
<td>4</td>
<td>0.1630</td>
</tr>
<tr>
<td>CAT2_EF</td>
<td>3.5384</td>
<td>2</td>
<td>0.1705</td>
</tr>
<tr>
<td>INTER</td>
<td>6.0114</td>
<td>8</td>
<td>0.6460</td>
</tr>
</tbody>
</table>

The above output indicates there is no difference of failure rate over the 5 sites, no difference of failure rate over the 3 levels of treatment, and no difference in failure rates between any two sites for each treatment condition, while controlling for the employment status and age of the patient at the significance level of 0.05.