Paper 208-26

A MACRO PROGRAM FOR ANOVA OR ANCONVA, USING PROC GLM OR PROC MIXED

Zaizai Lu, Kendle International Inc., Cincinnati, OH
David Shen, University of Cincinnati, Cincinnati, OH

ABSTRACT

Two-way Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA) are the two most commonly used statistical analysis procedures for continuous variables in clinical trial projects. The most common SAS® procedures to conduct ANOVA or ANCOVA are PROC GLM and PROC MIXED models. The paper introduces a macro program which enables the user to choose between PROC GLM and PROC MIXED, between ANOVA and ANCOVA, between CONTRAST statements and ESTIMATE statements. The program also gives the user a choice between an option with LSMEANS and without LSMEANS, with CONTRAST statements and without CONTRAST statements.

INTRODUCTION

Two-way Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA) are the two most commonly used statistical analysis procedures for continuous variables in clinical trial projects. We use these two tests almost every day, but there are no existing efficient macro programs to do these tests, especially a macro program with optional CONTRAST statements. The paper introduces a macro program we created which enables the user to choose between PROC GLM and PROC MIXED, between ANOVA and ANCOVA, between CONTRAST statements and ESTIMATE statements. The program also gives the user a choice between an option with LSMEANS and without LSMEANS, with CONTRAST statements and without CONTRAST statements.

This paper is composed of 5 parts. Part 1 includes the abstract and introduction, part 2 describes the data file used in the paper to generate the sample output of the SAS macro program. Part 3 describes the macro program in more details. Part 4 exhibits some examples of the applications of the macro program and their output. The final part is the conclusion of the paper.

DATA FILE

Throughout the paper, we used a data set called ONE, which contains 488 patients and seven variables: Patient ID, Investigation center, Treatment group, Patient's gender, Patient's race, Baseline pain value, and Change from the baseline pain value to the final visit. A more detailed description of the variables in the data set is as follows:

Variable	Type	Label
Patient	Char	Patient ID
Treat	Num	Treatment group
Center	Char	Investigation center
Base	Num	Baseline pain value
Change	Num	Change of pain from the baseline to final visit
Gender	Num	Patient's gender
Race	Num	Patient's race

MORE DETAILS OF THE MACRO

We named the macro as TESTED. There are altogether nine macro variables in this macro.

- Macro variable INDATA tells the user to supply the name of the data set to be analyzed, which uses conventional SAS® data set names.
- Macro variable TEST enables the user to select either GLM or MIXED test.
- Macro variable CLSLIST enables the user to provide the class variables to break down the analysis. There is no restriction on the number of class variables one supplies to a specific model. Different variables are separated by a space.
- Macro variable DVLIST tells the user to specify the dependent variable he/she wants to analyze. Only one dependent variable can be analyzed at one time.
- Macro variable IVLIST enables the user to put as many independent variables as he/she wants, thus resulting in two-way ANOVA, ANCOVA, or multiple-way analysis of variances. Interaction terms are allowed in both GLM and MIXED models. The user simply multiplies the two variables together and puts it in the model to create an interaction effect.
- Macro variable TRT_VAR looks for a treatment group variable, which will be used in the contrast statements or estimate statements for the treatment effect.
- Macro variable CONTRAST enables the user to select between CONTRAST statements and ESTIMATE statements. The available choices are CONTRAST, ESTIMATE. The default option is no CONTRAST or ESTIMATE statements.

- Macro variable LSTEST asks the user whether he/she wants Least-squares means test or not. If the user wants LSMEANS test, then he/she puts YES there, otherwise he can put a NO or simply leave it blank. The default option of the macro is no LSMEANS test.
- Macro variable OUTDATA tells the user to supply the name of the output data set to be produced, which uses conventional SAS® data set names. If the user only requests a simple GLM or MIXED model, the result output will only include the regular ANALYSIS OF VARIANCE or ANALYSIS OF COVARIANCE table. If the user requests CONTRAST statements, the result output will include the output for the CONTRAST statements as well as the regular ANOVA or ANCOVA table.

SOME EXAMPLES OF THE MACRO

Example 1: Single dependent variable for a simple GLM model without CONTRAST or LSMEANS statements.

```
%tested (indata=one,
         test=glm,
         dvlist=change,
         clslist=center treat,
         ivlist=center treat ,
         trt_var=treat,
         contrast=,
         lstest=,
         outdata=glmout1);
```

Figure 1. Example 1 output.

NAME	_SOURCE_	_TYPE_	DF	SS	F	PROB
CHANGE	ERROR	ERROR	443	87.0518	.	.
CHANGE	CENTER	SS1	38	11.1166	1.4887	0.034052
CHANGE	TREAT	SS1	2	10.6575	27.1175	0.000000
CHANGE	CENTER	SS3	38	10.9823	1.4707	0.038554
CHANGE	TREAT	SS3	2	10.6575	27.1175	0.000000

This example uses the PROC GLM procedure. It uses the data set ONE. The GLM model has one dependent variable (CHANGE), two independent variables (CENTER and TREAT), both of which are class variables. The name of the treatment variable is TREAT. It does not request CONTRAST statements or Least-squares means test. The options for the CONTRAST statements and LSMEANS test are left blank, indicating the default (no CONTRAST statements or LSMEANS test) will be selected. The resulting output is the Analysis of Variance (ANOVA) table and saved in a data set GLMOUT1.

Example 2: Single dependent variable for a GLM model with CONTRAST statements

```
%tested (indata=one,
         test=glm,
         dvlist=change,
         clslist=center treat,
         ivlist=center treat base,
         trt_var=treat,
         contrast=contrast,
         lstest=,
         outdata=glmout2);
```

Figure 2. Example 2 output.

NAME	_SOURCE_	_TYPE_	DF	SS	F	PROB
CHANGE	ERROR	ERROR	442	72.8149	.	.
CHANGE	CENTER	SS1	38	11.1166	1.7758	0.003827
CHANGE	TREAT	SS1	2	10.6575	32.3464	0.000000
CHANGE	BASE	SS1	1	14.2369	86.4207	0.000000
CHANGE	CENTER	SS3	38	11.3393	1.8114	0.002852
CHANGE	TREAT	SS3	2	10.0823	30.6006	0.000000
CHANGE	BASE	SS3	1	14.2369	86.4207	0.000000
CHANGE	TRT2 VS TRT1	CONTRAST	1	4.9391	29.9811	0.000000
CHANGE	TRT3 VS TRT1	CONTRAST	1	9.4295	57.2388	0.000000
CHANGE	TRT3 VS TRT2	CONTRAST	1	0.6820	4.1399	0.042479

Example 2 also uses the PROC GLM procedure. It is very similar to Example One, except that in this example, we requested the CONTRAST statements, and added the variable BASE as a covariate. The resulting output includes the CONTRAST statements in addition to the Analysis of Covariance (ANCOVA) table.

Example 3: Single dependent variable for a GLM model with CONTRAST and LSMEANS statements.

```
%tested (indata=one,
         test=glm,
         dvlist=change ,
         clslist=center treat,
         ivlist=center treat ,
         trt_var=treat,
         contrast=contrast,
         lstest=yes,
         outdata=glmout3);
```

Figure 3. Example 3 output.

NAME	_SOURCE_	_TYPE_	DF	SS	F	PROB
CHANGE	ERROR	ERROR	443	87.0518	.	.
CHANGE	CENTER	SS1	38	11.1166	1.4887	0.03405
CHANGE	TREAT	SS1	2	10.6575	27.1175	0.00000
CHANGE	CENTER	SS3	38	10.9823	1.4707	0.03855
CHANGE	TREAT	SS3	2	10.6575	27.1175	0.00000
CHANGE	TRT2 VS TRT1	CONTRAST	1	6.0844	30.9632	0.00000
CHANGE	TRT3 VS TRT1	CONTRAST	1	9.4883	48.2854	0.00000
CHANGE	TRT3 VS TRT2	CONTRAST	1	0.3569	1.8165	0.17842

NAME	TREAT	LSMEAN	STDERR
CHANGE	1	-0.35825	0.035738
CHANGE	2	-0.63406	0.036266
CHANGE	3	-0.70096	0.035753

Example 3 also uses the PROC GLM procedure. It is very similar to the previous two examples, except that in this example, we requested the LSMEANS test in addition to the CONTRAST statements. The resulting output includes the LSMEANS output in addition to the regular ANOVA table and the CONTRAST output.

Example 4: Single dependent variable for a GLM model with interaction terms.

```
%tested (indata=one,
         test=glm,
         dvlist=change,
         clslist=center treat gender
                 race,
         ivlist=center treat gender
                 race gender*race,
         trt_var=treat,
         contrast=,
         lstest=,
         outdata=glmout4);
```

Figure 4. Example 4 output.

NAME	_SOURCE_	_TYPE_	DF	SS	F	PROB
CHANGE	ERROR	ERROR	434	83.9090	.	.
CHANGE	CENTER	SS1	38	11.1166	1.5131	0.02884
CHANGE	TREAT	SS1	2	10.6575	27.5616	0.00000
CHANGE	GENDER	SS1	1	0.1127	0.5828	0.44564
CHANGE	RACE	SS1	5	2.4191	2.5024	0.02997
CHANGE	GENDER*RACE	SS1	3	0.6110	1.0535	0.36869
CHANGE	CENTER	SS3	38	11.9330	1.6242	0.01277
CHANGE	TREAT	SS3	2	10.1158	26.1609	0.00000
CHANGE	GENDER	SS3	1	0.4892	2.5305	0.11239
CHANGE	RACE	SS3	5	1.8555	1.9194	0.08986
CHANGE	GENDER*RACE	SS3	3	0.6110	1.0535	0.36869

In this example, we added an interaction term GENDER*RACE to the GLM model. Since we did not request CONTRAST or LSMEANS statements, the result is also a regular ANCOVA table.

Example 5: Single dependent variable for a simple MIXED model without CONTRAST or LSMEANS statements.

```
%tested (indata=one,
         test=mixed,
         dvlist=change,
         clslist=center treat,
         ivlist=center treat base,
         trt_var=treat,
         contrast=,
         lstest=,
         outdata=mixout1);
```

Figure 5. Example 5 output.

SOURCE	NDF	DDF	F	P_F
CENTER	38	442	1.81	0.0029
TREAT	2	442	30.60	0.0001
BASE	1	442	86.42	0.0001

This example uses the PROC MIXED procedure. It uses the data set ONE. The MIXED model has one dependent variable (CHANGE), three independent variables, two of which are class variables (CENTER and TREAT). The name of the treatment variable is TREAT. It does not request contrast statements or least-squares means test. The options for the CONTRAST/ESTIMATE statements and LSMEANS test are left blank, indicating the default (no CONTRAST/ESTIMATE statements or LSMEANS test) will be selected. The resulting output is saved in a data set MIXOUT1.

Example 6: Single dependent variable for a MIXED model with ESTIMATE statement.

```
%tested (indata=one,
         test=mixed,
         dvlist=change,
         clslist=center treat,
         ivlist=center treat ,
         trt_var=treat,
         contrast=estimate,
         lstest=,
         outdata=mixout2);
```

Figure 6. Example 6 output.

SOURCE	NDF	DDF	F	P_F
CENTER	38	443	1.47	0.0386
TREAT	2	443	27.12	0.0001

PARM	EST	SE	DF	T	P_T
TRT2 VS TRT1	-0.27580980	0.04956630	443	-5.56	0.0001
TRT3 VS TRT1	-0.34271598	0.04932037	443	-6.95	0.0001
TRT3 VS TRT2	-0.06690617	0.04964216	443	-1.35	0.1784

Example 6 uses the PROC MIXED procedure. In this example, we requested the ESTIMATE statements. The resulting output includes the ESTIMATE statements in addition to the regular ANOVA table.

Example 7: Single dependent variable for a simple MIXED model with CONTRAST and LSMEANS statements.

```
%tested (indata=one,
         test=mixed,
         dvlist=change,
```

```
clslist=center treat,
ivlist=center treat ,
trt_var=treat,
contrast=contrast,
lstest=yes,
outdata=mixout3);
```

Figure 7. Example 7 output.

SOURCE	NDF	DDF	F	P_F
CENTER	38	443	1.47	0.0386
TREAT	2	443	27.12	0.0001
TRT2 VS TRT1	1	443	30.96	0.0001
TRT3 VS TRT1	1	443	48.29	0.0001
TRT3 VS TRT2	1	443	1.82	0.1784

EFFECT	TREAT	_LSMEAN_	_SE_	_DF_	_T_	_PT_
TREAT	1	-0.35824774	0.03573761	443	-10.02	0.0001
TREAT	2	-0.63405755	0.03626553	443	-17.48	0.0001
TREAT	3	-0.70096372	0.03575307	443	-19.61	0.0001

In Example 7, we used PROC MIXED procedure, and requested CONTRAST statements and LSMEANS test instead of ESTIMATE statements. The resulting output is shown in Figure 7.

Example 8: Single dependent variable for a MIXED model with interaction terms and LSMEANS statements.

```
%tested (indata=one,
         test=mixed,
         dvlist=change,
         clslist=center treat,
         ivlist=center treat
                center*treat,
         trt_var=treat,
         contrast= ,
         lstest=yes,
         outdata=mixout4);
```

Figure 8. Example 8 output.

SOURCE	NDF	DDF	F	P_F
CENTER	38	367	1.49	0.0342
TREAT	2	367	32.84	0.0001
CENTER*TREAT	76	367	1.07	0.3427

EFFECT	TREAT	_LSMEAN_	_SE_	_DF_	_T_	_PT_
TREAT	1	-0.31598957	0.03788066	367	-8.34	0.0001
TREAT	2	-0.62548563	0.03878115	367	-16.13	0.0001
TREAT	3	-0.73251748	0.03734297	367	-19.62	0.0001

In this example, we used PROC MIXED procedure with an interaction term, and requested LSMEANS test. The bottom part is the LSMEANS test result.

CONCLUSION

This paper demonstrates the use of a macro program we created to conduct PROC GLM or PROC MIXED tests. Although the results from GLM and MIXED procedures are not exactly numerically the same, they are very close. Our macro program has several advantages: 1) It enables the user to choose between PROC GLM and PROC MIXED, depending on individual's preference or study requirement. 2) It can automatically count the total number of treatment groups in the analysis data set. 3) It automatically sets up the contrast statements based on the number of treatment groups in the data set. 4) It can include as many class variables as one wants. 5) It can have as many independent variables as one wants. 6) The CONTRAST statements can handle as many as 10 treatment groups. 7) All the macro variables are not case-sensitive. They accept both upper case and lower case.

Most important of all, the macro is easy to use and understand.

REFERENCES

SAS Institute Inc. (1990) SAS/STAT® User's Guide, Version 6, Fourth Edition, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990) SAS® Guide to Macro Processing, Version 6, Fourth Edition, Cary, NC: SAS Institute Inc.

TRADEMARKS

CONTACT INFORMATION

For additional information or source code of the SAS macro program, contact the author at:

Zaizai Lu
Kendle International Inc.
1200 Carew Tower
441 Vine Street
Cincinnati, OH 45202
lu.zaizai@kendle.com