

## **%QPCR\_MIXED: SAS® macro to elicit and fit linear mixed models to qPCR data.**

Juan P. Steibel

### **Description**

%QPCR\_MIXED can be used to generate SAS PROC MIXED statements to analyze qPCR data. It is intended to facilitate implementation of the method proposed in Steibel JP, Poletto R, Coussens PM and Rosa GJM 2009 (A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data, under revision).

At present QPCR\_MIXED can generate code to analyze the following basic experimental designs:

- 1) Randomized complete block design with one factor.
- 2) Randomized complete block design with two factors including.
- 3) Randomized complete block design with two factors in split-plots.
- 4) Completely randomized design with one factor.
- 5) Completely randomized design with two factors including.
- 6) Completely randomized design with two factors in split-plots.

These six basic designs span a number of experimental situations used in practice. For example: time-course experiments can be analyzed as split-plot designs. Similarly, three-way factorial can be collapsed into a two-way factorial by having one of the factors represent a two-way interaction.

At present a single control gene is permitted, but this requirement will be relaxed in future versions.

If your data does not match the described designs and you want to use %QPCR\_MIXED for your analysis, please email the author with an inquiry as some other designs could be accommodated in a near future. You may also want to check the “to-do” list in the website to see which options are currently under development.

### **Support**

This macro is written and supported by Juan P. Steibel at Michigan State University. Please email him ([steibelj@msu.edu](mailto:steibelj@msu.edu)) with suggestions, bug-reports and questions.

### **Citation**

QPCR\_MIXED is provided for free use. If you use all or part of the macro code, please cite the following paper:

*Steibel JP, Poletto R, Coussens PM and Rosa GJM 2009. A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. Genomics. Accepted.*

### **Input parameters:**

The input parameters of QPCR\_MIXED are:

```
%MACRO QPCR_MIXED(indata=my_data, response=ct,  
                    gene_var=detector, control_gene=BetaActi,  
                    fixed1=factor_1, fixed2=, rep_type=RAND,  
                    rep_id=unit, design_structure=NONE, cont=NO,
```

### **Indata**

Name of the datafile. It should be an existing SAS data file that contains one row per amplification reaction (typically this is a read from a single well in the thermocycler). It should also include the following columns (variables):

- 1) response variable (a quantity proportional to initial log-mRNA concentration, i.e CT). Missing values should be represented by the “dot character” (.).
- 2) gene variable: this column contains the names of the genes, at least two genes: a test and a control gene should be present.
- 3) Fixed effects factor: one or two columns indicating fixed effects factors
- 4) Experimental or replication unit: a column indicating the biological replication or the block (if a block design).

*Note: if you do not know how to create this file in SAS, see the [examples section of this guide](#).*

### **response**

Name of the response variable in **indata**. By default: ct.

### **gene\_var**

The name of the gene variable in **indata**.

**control\_gene:** name of the control gene in **gene\_var**. This should match exactly the name in your file, otherwise an error will be produced.

**fixed1:** name of the first factor of fixed effects.

**fixed2:** name of the second factor of fixed effects. OPTIONAL. If the experiment involves only one factor, do not specify this parameter.

**rep\_type:** either RAND or BLOCK. If experimental units are arranged in blocks, set this variable as BLOCK, if experimental units are arranged completely at random, set this variable to RAND. If your design is a split plot design, the experimental unit is assumed to be the main plot factor unit and it can be specified as RAND or BLOCK using this parameter.

**rep\_id:** name of the variable identifying the experimental units or blocks.

**design\_structure:** either NONE or SPLIT. If the design is a two-way factorial in a split-plot arrangement, set this parameter to SPLIT otherwise keep the default value: NONE. This parameter can be combined with **rep\_type** to generate split-plot designs with main plot factor arranged in blocks.

**cont:** YES/NO. if **cont** is set to YES, %QPCR\_MIXED invokes a sub-macro (%QPCR\_CONTRAST) that will generate code to obtain estimates and contrast of interest. The submacro should be declared before setting this option to YES.

**df:** any valid method to obtain denominator degrees of freedom in SAS PROC MIXED (SATTERTH, CONTAIN, KR, etc). By default this parameter is set to SATTERTH and it should

not be modified. But in some cases, the SATTERTH option may take too long and the user can try with the CONTAIN option to make sure that the generated code runs properly before running the final job with the computationally more expensive **df=SATTERTH** option.

**file:** PRINT/RUN. If **file=PRINT**, the SAS PROC MIXED statements will be printed to the standard output. The printed statements can be copied and pasted into the SAS editor and checked for consistency before running them. This is the recommended way to use this macro. However, the option **file=RUN** is provided too and it can be used to generate statements and run them in one stage. This is only recommended for routine analysis of previously tested cases.

## Examples

These and other examples are available as SAS code from the %QPCR\_MIXED website. See below for a step-by-step tutorial using your own data.

### 1) One way ANOVA in completely randomized design.

The data were obtained from [1].

Part of datafile:

Obs	Obs	File	Well	Sample	Gene	Ct	Pipe	Log2Amt	Frequency	mut
	1	A0182094	Well1349	CLL30	18S	8.4511	Pipe7	29.7301	9	1
	2	A0182094	Well1350	CLL30	18S	8.1998	Pipe7	29.9813	9	1
	3	A0182094	Well1373	CLL30	18S	8.4485	Pipe7	29.7327	9	1
	4	A0182094	Well1374	CLL30	18S	8.3740	Pipe7	29.8072	9	1
	5	A0182106	Well1349	CLL35	18S	8.7024	Pipe7	29.4787	9	1
	6	A0182106	Well1350	CLL35	18S	8.5759	Pipe7	29.6053	9	1
...										
2198	2198	A0182114	Well1228	CLL61	iNos(Nos	33.2626	Pipe4	4.9185	8	1

The preferred response variable in this dataset is Log2Amt because it represents a CT corrected for amplification efficiency. There is only one experimental factor: mutant status: (mut=1/0). The dataset comprises 64

Running the statement:

```
%QPCR_MIXED(indata=analysis, response=Log2Amt, gene_var=gene, control_gene=18S,  
            fixed2=, rep_id=sample, rep_type=RAND, design_structure=NONE,  
            cont=NO, file=PRINT)
```

Produces:

```
/*SAS PROC MIXED STATEMENTS*/  
PROC MIXED data=analysis ;  
Class gene mut sample;  
model Log2Amt = gene*mut/noint solution;  
random sample(mut);  
random sample(mut)/group=gene;  
run;
```

### 2) two-way factorial in a randomized complete block design.

This is part of a larger dataset published by [2]

Part of datafile:

Obs	Well	Type	Name	group	Detector	Ct	Std Dev
-----	------	------	------	-------	----------	----	------------

## %QPCR\_MIXED user guide

```
1   A1   Clinical   2437   NIL   BetaActi   15.71   0.101
2   B1   Clinical   2437   NIL   BetaActi   15.57   0.101
3   A2   Clinical   2437   PTB   BetaActi   17.10   0.016
4   B2   Clinical   2437   PTB   BetaActi   17.12   0.016
5   G3   Clinical   3361   NIL   BetaActi   26.28   0.238
...
96  F2   Clinical   7865   PTB   IL-12p35   31.23   0.180
```

ct is the response variable in this case. There are two experimental factors: Type (Clinical, Subclinical and Control) refers to paratuberculosis status of each cow, Group (NIL/PTB) refers to PTB challenge (or not) of blood cells from each cow. The experimental unit is the cow (indicated by Name). The resulting design is a split-plot design, with Type being the main-plot factor and Group being the sub-plot factor.

```
%QPCR_MIXED(indata=ptb, response=ct, gene_var=detector,
             control_gene=BetaActi, fixed1=type, fixed2=group,
             rep_id=Name, rep_type=RAND, design_structure=SPLIT,
             cont=YES, df=SATTERTH, file=PRINT);
```

### Output:

```
/*SAS PROC MIXED STATEMENTS*/
PROC MIXED data=ptb ;
Class detector type group Name;
model ct = detector*type*group/noint ddfm=SATTERTH solution;
random Name(type) group*Name(type);
random Name(type) group*Name(type)/group=detector;

/*CONTRASTS AND ESTIMATES NOT PRINTED IN THIS USED GUIDE*/

run;
```

### 3) Running example 2):

To actually fit the mixed model from example 2) we can copy and paste the generated code into SAS editor and hit the RUN button (recommended). Or we can re-run the macro as follows:

```
%QPCR_MIXED(indata=ptb, response=ct, gene_var=detector,
             control_gene=BetaActi, fixed1=type, fixed2=group,
             rep_id=Name, rep_type=RAND, design_structure=SPLIT,
             cont=YES, df=SATTERTH, file=RUN);
```

### Selected output:

#### a) Proc mixed ANOVA table:

Type 3 Tests of Fixed Effects				
Effect	Num	Den	F Value	Pr > F
	DF	DF		
Detector*Type*group	12	12.4	675.91	<.0001

This piece of output is of little relevance in our analysis because we are interested in inferences on type and group factors once the control gene correction has been introduced.

#### b) Gene-specific ANOVA table

## %QPCR\_MIXED user guide

Label	Contrasts		F Value	Pr > F
	Num	Den		
type geneIL-12P35	2	13	1.96	0.1805
group geneIL-12P35	1	9.96	3.74	0.0822
type*group geneIL-12P35	5	15.8	1.71	0.1889

The actual ANOVA table that we want is under the CONTRASTS section of the output. If the dataset comprise more than one gene, this table will include many rows to accommodate all test genes. The control gene correction has been introduced before obtaining this table.

In this particular case, there is no evidence of interaction between type and group and only a marginally significant effect of group in the normalized expression of IL12. We should stop here and conclude that there is no evidence in favor of significant differential expression of IL12 normalized to the expression of 18S ribosomal protein (control gene).

### c) Estimates

Label	Estimate	Standard	DF	t Value	Pr >  t
		Error			
type:Clinical-Control gene IL-12P35	3.2194	2.3805	9.32	1.35	0.2082
group:NIL-PTB geneIL-12P35	1.0392	0.5377	9.96	1.93	0.0822
type:Control-Subclini group=PTB gene:IL-12P35	-5.0013	2.4699	10.7	-2.02	0.0685
group:NIL-PTB type=Clinical gene:IL-12P35	1.5912	0.9313	9.96	1.71	0.1184

/\*TRUNCATED OUTPUT FROM SAS\*/

log-fold change estimates are provided under the ESTIMATES section of the SAS output. This table includes the estimated differences in CT of IL12 normalized to 18S, for pair-wise contrast of interest. The estimate column in this table is comparable to the  $\Delta\Delta\text{CT}$  [3]. No difference was significant ( $\alpha=0.05$ ), as expected from the ANOVA table.

### 4) Using your own data files from an excel spreadsheet

In order to use %QPCR\_MIXED with your own data, you will need to upload your data into SAS following the steps described in this section.

- Format your data using MS Excel. The structure of the datafile is described in page 2 of this guide ([description of indata file](#)). Make sure to use search and replace to substitute missing values with ".". Save file in excel format to your disk.
- Upload file into sas using the following statements:

```
proc import
datafile="C:\your_path\file_name.xls"
OUT=my_data
DBMS=EXCEL replace;
RUN;
PROC PRINT DATA=my_data;
run;
```

my\_data and "C:\your\_path\file\_name.xls" are the names of the (new) SAS dataset and the excel file respectively.

- Check the import: If data import was completed successfully, you should see your data printed in the standard SAS output.
- Run %QPCR\_MIXED as with previous examples.

### **Warning on using this macro**

This macro is no replacement for expert statistical advice. It may be tempting to input the data, specify the design parameters and let the macro automatically run the analysis for you. But we recommend that the code generated by the macro is revised and checked for errors. Linear mixed models can encompass a multitude of different variants and can be used to fit data from any experimental designs.

There may be some situations where %QPCR\_MIXED will produce a result that is suboptimal compared to alternative linear models. The fact that %QPCR\_MIXED runs without error is no warranty that the analysis is the most appropriate for a particular dataset.

Checking and comparing results with other methods is highly recommended.

### REFERENCES

- [1] Abruzzo L. V., Lee K. Y., Fuller A., Silverman A., Keating M. J., Medeiros L. J., and Coombes K. R., Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data, *Biotechniques* 38 (2005) 785-92.
- [2] Coussens P. M., Colvin C. J., Rosa G. J., Perez Laspiur J., and Elftman M. D., Evidence for a novel gene expression program in peripheral blood mononuclear cells from *Mycobacterium avium* subsp. paratuberculosis-infected cattle, *Infect Immun* 71 (2003) 6487-98.
- [3] Livak K. J., and Schmittgen T. D., Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method, *Methods* 25 (2001) 402-8.