Repeated measures ANOVA using the mixed models in XLSTAT

demoANOrep.xls

Dataset for repeated measures ANOVA using the mixed models

An Excel sheet with both the data and the results can be downloaded by clicking here.

The data correspond to an experiment in which a treatment for depression is studied. Two groups of patients (1: control / 2: treatment) have been followed at five different times (0: pre-test, 1: one month post-test, 3: 3 months follow-up and 6: 6 months follow-up). The dependant variable is a depression score.

We have performed a repeated measures ANOVA in order to determine the effect of the treatment and the effect of time on the depression score. The repeated measures ANOVA model is the same as the classical ANOVA model with interactions:

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk}$$

We have two fixed factors (time and group) and one interaction factor (time*group). The difference between classical ANOVA and repeated measures ANOVA is that measures on the same patient at different times are not supposed to be independent and, thus, the covariance matrix of e is not diagonal.

XLSTAT uses the mixed models theory to treat repeated measures ANOVA and this raises some differences. Some supplementary options are available like the choice between many covariance structures for the covariance matrix of the error term. Throughout this tutorial we will use the compound symmetry structure. Please consult the XLSTAT help for more details on covariance structures.

Data structure for repeated measures ANOVA

Data should have a particular shape when using the mixed model methodology for repeated measures ANOVA. All the measures have to be on the same column with:

- a factor called repeated factor which indicates what repetition is associated to this measure
- a factor called subject factor which indicates what subject is associated to this measure.

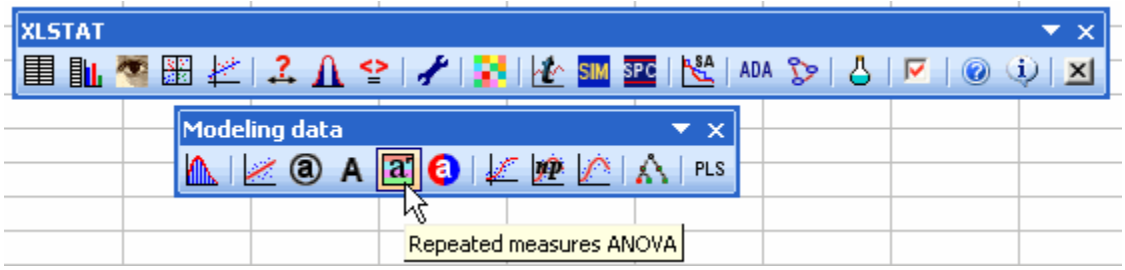
In our example, each patient will appear in 4 different lines. The data has to be in the following form:

subj	time	group	dv
1	0	1	296
1	1	1	175
1	3	1	187
1	6	1	242
2	0	1	376
2	1	1	329
2	3	1	236
2	6	1	126
3	0	1	309
3	1	1	238
3	3	1	150
3	6	1	173
4	0	1	222
4	1	1	60
4	3	1	82
4	6	1	135

If your data is in a different form, one column for each measure, you should transform it in order to obtain the preceding structure.

Setting up the repeated measures ANOVA using the mixed models

After opening XLSTAT, select the **XLSTAT / Modeling data / Repeated measures ANOVA** command, or click on the corresponding button of the **Modeling data** toolbar (see below).



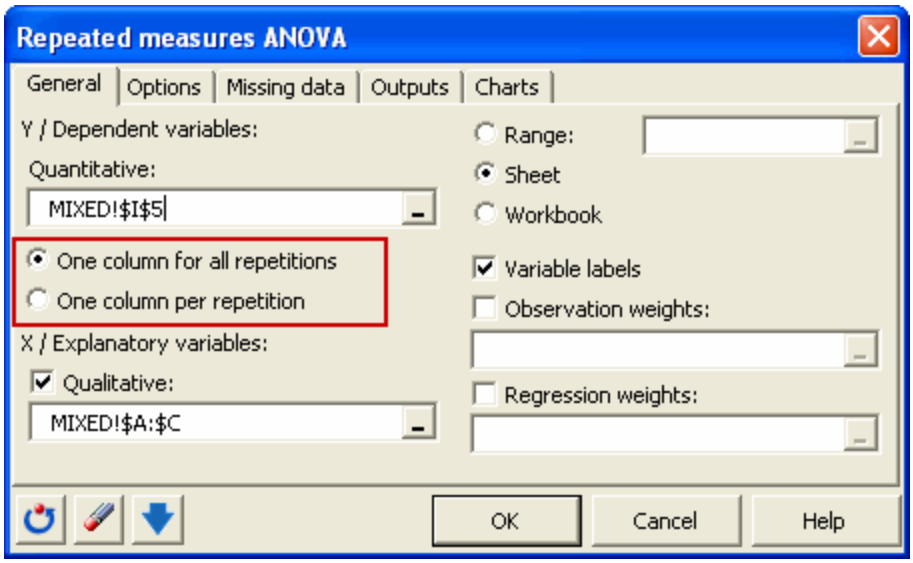
Once you've clicked on the button, the **repeated measures ANOVA** dialog box appears. Select the data on the Excel sheet.

The **Dependent variable** (or variable to model) is here the "dv".

Our aim is to determine the effect of the group, the time and the interaction between the two on the variability of the depression score.

As we selected the column title for the variables, we left the option **Variable labels** activated.

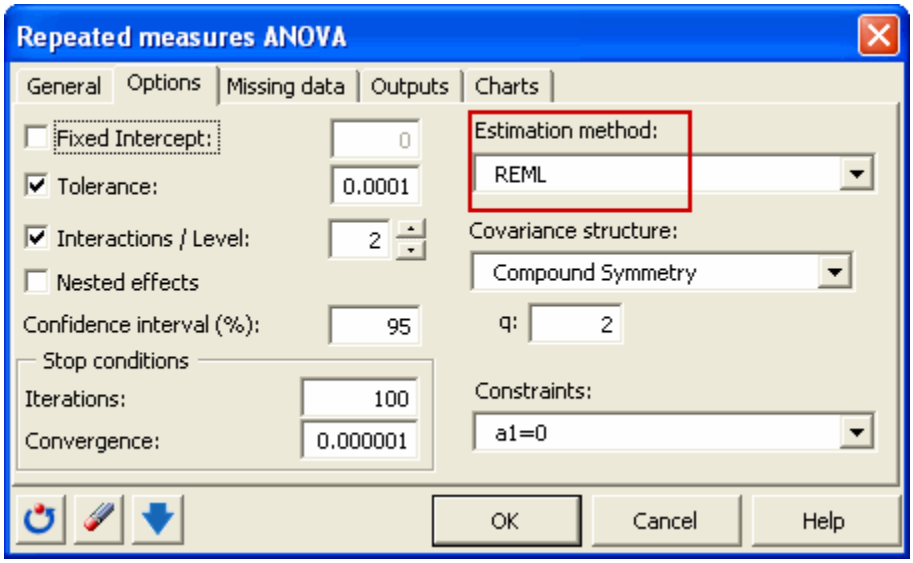
"Repeated" and "subject" factors must be selected as **explanatory variables**.



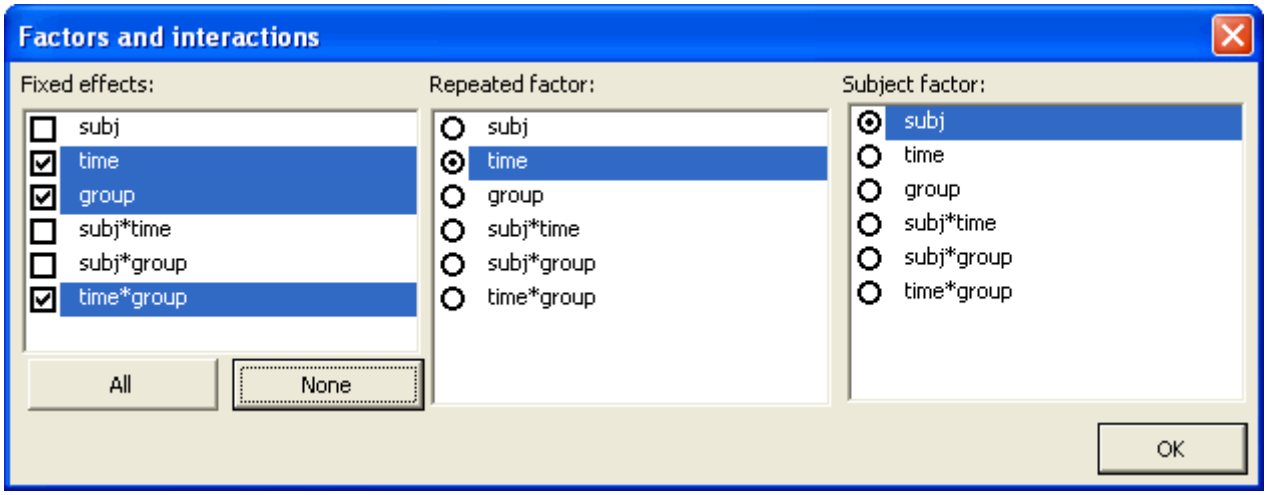
The **interactions** option is activated on the options tab, and the **maximum level of interaction** is set to 2.

We left the constraint option at **a1=0**, meaning that we want the model to be built on the assumption that the control group has the standard effect on the score. Although you have to apply a constraint to the model in ANOVA for theoretical reasons, it will not affect the results (goodness of fit). The only difference it makes is in the actual writing of the model.

The **covariance structure** selected is the default one which is **compound symmetry**.



Once you have clicked on the **OK** button, a dialog box is displayed so that you can choose which factors have to be taken into account in the model. The **fixed effects** are "time", "group" and "time*group", the **repeated factor** is "time" and the **subject** factor is "subj".



Note: A factor cannot be the subject factor and a fixed effect at he same time. Repeated and subject factors have to be different and both qualitative.

Once you have clicked on the **OK** button, the computation starts. The results will then be displayed.

Interpreting the results of a repeated measures ANOVA using the mixed models

The first results displayed by XLSTAT are the goodness of fit coefficients.

Goodness of fit statistics:	
Observations	96
Sum of weights	96
-2 Res Log(Likelihood)	1001,258
AIC	1005,258
AICC	1005,387
SBC	1010,387
CAIC	1012,387
Iterations	6
Covariance parameters	2
Number of fixed effects	3
Number of random effects	0
Number of subjects	24
Maximum number of observe	4

Model parameters are obtained using the restricted maximum likelihood (REML) method and

will be different as when a classical ANOVA model is applied. All indexes are used to compare models with different covariance structures.

Covariance parameters are displayed on the next table with their associated Z test. We can see that both parameters are significant.

Covariance parameters - Repeated factor:

	Parameters	Standard error	Z	Pr > Z
Variance	2781,872	484,338	5,744	< 0,0001
Sigma 1	2526,293	978,817	2,581	0,010

Model implied covariance and correlation matrix can as well be displayed.

The null model likelihood ratio test compares the model obtained with the specified covariance structure and the model obtained with the classical diagonal covariance structure. The obtained p-value can be used to assess the significance of the model fit.

Null model likelihood ratio test:

DF	Chi-square	Pr > Chi²
1	23,131	< 0,0001

We can see that choosing the compound symmetry structure has a positive effect on the model fit.

The type III tests of fixed effects show that all factors have significant effects on the depression score. These test are very important, they replace the Type III SS of the classical ANOVA model.

Type III tests of fixed effects:

Effects	Num DF	Den DF	F	Pr > F
time	3	88	37,615	< 0,0001
group	1	88	17,886	< 0,0001
time*group	3	88	7,164	0,000

From the results displayed in the Type III table, we can see that the "Time" variable is the one that has the highest impact on the model. When we look at the model parameters (see below), we can see that time 1, 3 and 6 have a negative impact on the depression score. Patients are less depressed as time passes. Being in the treatment group has also a negative impact on the depression score.

Model parameters:						
Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	304,333	21,032	14,470	< 0,0001	262,585	346,082
time-0	0,000					
time-1	-119,333	30,451	-3,919	0,000	-179,779	-58,888
time-3	-113,500	30,451	-3,727	0,000	-173,946	-53,054
time-6	-107,167	21,532	-4,977	< 0,0001	-149,908	-64,425
group-1	0,000					
group-2	-90,500	29,744	-3,043	0,003	-149,541	-31,459
time-0*gro	0,000					
time-0*gro	66,583	30,451	2,187	0,031	6,138	127,029
time-1*gro	71,667	30,451	2,353	0,021	11,221	132,112
time-1*gro	0,000	0,000				
time-3*gro	24,917	30,451	0,818	0,415	-35,529	85,362
time-3*gro	0,000	0,000				
time-6*gro	0,000	0,000				
time-6*gro	0,000	0,000				

The table depicted above can be used to analyze the impact of the explanatory variables on the depression score and/or to predict the average yield in a situation not yet covered by the experiment, such as the time 6 with the treatment group. In this particular example, the average depression score would be 106.7, given the fact that the influence of the interaction is unknown.

This study has shown that both time and treatment have both a negative significant impact on the depression score.

Some other output can be useful and are available in XLSTAT like residuals, residuals charts, least square means charts…