**Proc Mixed for ANOVA Designs Including Correlated Samples**

Please read [David Howell’s document](https://www.uvm.edu/~dhowell/StatPages/Missing_Data/Mixed%20Models%20for%20Repeated%20Measures.pdf) on using Proc Mixed. Both the data and the SAS code used here came from his site. We have data comparing the effectiveness of an experimental method for treating depression to that of a placebo group. For each subject we have a measure of depression at baseline, post-test, three month follow-up, and six month follow-up. When you look at the data file, you will find it is in what I call “the univariate setup.”

**Data** Howell; infile 'C:\Users\Vati\Desktop\WicksellLong.dat';

Input subject time group dv;

Label Group = 'Treatment vs Control'; **run**;

|  |
| --- |
| **The Proc Mixed analysis assuming compound symmetry.** |

**Proc** **Mixed**; class group subject time;

model dv = group time group\*time;

repeated time /subject = subject type = cs rcorr; **run**;

* Notice that group, subject, and time are all declared classification variables.
* “type = cs” instructs SAS to assume that there is compound symmetry (akin to sphericity) in the population.
* “rcorr” displays the estimated correlation matrix

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HOWELL |
| **Dependent Variable** | dv |
| **Covariance Structure** | Compound Symmetry |
| **Subject Effect** | subject |
| **Estimation Method** | REML |
| **Residual Variance Method** | Profile |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **group** | 2 | 1 2 |
| **subject** | 24 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 |
| **time** | 4 | 0 1 3 6 |

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 2 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 24 |
| **Max Obs Per Subject** | 4 |

| **Number of Observations** | |
| --- | --- |
| **Number of Observations Read** | 96 |
| **Number of Observations Used** | 96 |
| **Number of Observations Not Used** | 0 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 1024.25281678 |  |
| **1** | 1 | 1000.80478297 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

| **Estimated R Correlation Matrix for subject 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.4791 | 0.4791 | 0.4791 |
| **2** | 0.4791 | 1.0000 | 0.4791 | 0.4791 |
| **3** | 0.4791 | 0.4791 | 1.0000 | 0.4791 |
| **4** | 0.4791 | 0.4791 | 0.4791 | 1.0000 |

* Notice that the estimated correlation matrix (for the DV at the four times) is compound symmetric – the elements on the main diagonal are a constant (1) and the elements off the main diagonal are a constant (.4791).

| **Covariance Parameter Estimates** | | |
| --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** |
| **CS** | **subject** | 2539.36 |
| **Residual** |  | 2760.62 |

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 1000.8 |
| **AIC (smaller is better)** | 1004.8 |
| **AICC (smaller is better)** | 1004.9 |
| **BIC (smaller is better)** | 1007.2 |

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 1 | 23.45 | <.0001 |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **group** | 1 | 22 | 13.71 | 0.0012 |
| **time** | 3 | 66 | 45.14 | <.0001 |
| **group\*time** | 3 | 66 | 9.01 | <.0001 |

|  |
| --- |
| **Same analysis but specifying an unstructured covariance matrix.** |

**Proc** **Mixed**; class group subject time;

model dv = group time group\*time;

repeated time /subject = subject type = un hlm hlps rcorr; **run**;

* “type = un hlm hlps” tells SAS not to assume any particular correlation/covariance matrix, but rather to estimate it from the observed data.
* The “hlm” and “hlps” result the application of methods that do a better job of estimating the *df*.

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HOWELL |
| **Dependent Variable** | dv |
| **Covariance Structure** | Unstructured |
| **Subject Effect** | subject |
| **Estimation Method** | REML |
| **Residual Variance Method** | None |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 10 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 24 |
| **Max Obs Per Subject** | 4 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 1024.25281678 |  |
| **1** | 1 | 975.37524019 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

Baseline Post-test 3 Month 6 Month

| **Estimated R Correlation Matrix for subject 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.5695 | 0.5351 | -0.01683 |
| **2** | 0.5695 | 1.0000 | 0.8612 | 0.4456 |
| **3** | 0.5351 | 0.8612 | 1.0000 | 0.4202 |
| **4** | -0.01683 | 0.4456 | 0.4202 | 1.0000 |

* Notice that the off-diagonal correlations are not constant. Since these were estimated from the observed data, it is pretty clear that compound symmetry should not be assumed.

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 975.4 |
| **AIC (smaller is better)** | 995.4 |
| **AICC (smaller is better)** | 998.2 |
| **BIC (smaller is better)** | 1007.2 |

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 9 | 48.88 | <.0001 |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **group** | 1 | 22 | 13.71 | 0.0012 |
| **time** | 3 | 22 | 31.91 | <.0001 |
| **group\*time** | 3 | 22 | 10.08 | 0.0002 |

| **Type 3 Hotelling-Lawley-McKeon Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **time** | 3 | 20 | 29.01 | <.0001 |
| **group\*time** | 3 | 20 | 9.17 | 0.0005 |

| **Type 3 Hotelling-Lawley-Pillai-Samson Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **time** | 3 | 20 | 29.01 | <.0001 |
| **group\*time** | 3 | 20 | 9.17 | 0.0005 |

* When we assumed compound symmetry, our error df for the repeated effects were 66. Now they are 20. This loss of *df* is from estimating the structure of the variance/covariance matrix, and will result in the loss of power.

Look back at the estimated correlation matrix. Notice that times that are close to one another (for example, pretest correlates with post-test *r* = .57) are more strongly correlated than are times far from each other (for example, pretest correlates with post-test *r* = -0.02). We can save some *df* if we tell SAS to estimate the variance/covariance matrix assuming an autoregressive model. “AR(1)” would tell SAS to estimate the Time 1 to Time 3 correlation as the square of the correlation between Time 1 and Time 2, and the Time 1 to Time 4 correlation as the cube of the correlation between Time 1 and Time 2, and so on. Each correlation is based on the value of the preceding correlation. “AR(2) would tell SAS to base its estimate of each correlation from the values of the preceding two correlations.

|  |
| --- |
| Same analysis but specifying an autoregressive covariance matrix. |

**Proc** **Mixed**; class group subject time;

model dv = group time group\*time;

repeated time /subject = subject type = AR(**1**) rcorr; **run**;

AR(1) – as the times get further apart, their correlation decreases

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HOWELL |
| **Dependent Variable** | dv |
| **Covariance Structure** | Autoregressive |
| **Subject Effect** | subject |
| **Estimation Method** | REML |
| **Residual Variance Method** | Profile |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 2 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 24 |
| **Max Obs Per Subject** | 4 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 1024.25281678 |  |
| **1** | 2 | 991.55035479 | 0.00000108 |
| **2** | 1 | 991.54990599 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

| **Estimated R Correlation Matrix for subject 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.6112 | 0.3735 | 0.2283 |
| **2** | 0.6112 | 1.0000 | 0.6112 | 0.3735 |
| **3** | 0.3735 | 0.6112 | 1.0000 | 0.6112 |
| **4** | 0.2283 | 0.3735 | 0.6112 | 1.0000 |

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 991.5 |
| **AIC (smaller is better)** | 995.5 |
| **AICC (smaller is better)** | 995.7 |
| **BIC (smaller is better)** | 997.9 |

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 1 | 32.70 | <.0001 |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **group** | 1 | 22 | 14.20 | 0.0011 |
| **time** | 3 | 66 | 36.37 | <.0001 |
| **group\*time** | 3 | 66 | 11.66 | <.0001 |

**Missing Data**

With the traditional analysis (GLM), one cannot have any missing data. Of course, one can impute missing values. Proc Mixed can handle missing data without needing any imputation. Using a data file provided by Howell,

**Data** Howell; infile 'C:\Users\Vati\Desktop\WicksellLongMiss.dat';

Input subject group time dv;

Label Group = 'Treatment vs Control'

Time = 'Time of Measurement starting at 0'; **run**;

Title 'The Proc Mixed analysis with missing data and compound symmetry.';

**Proc** **Mixed**; class group subject time;

model dv = group time group\*time;

repeated time /subject = subject type = cs rcorr; **run**;

Title 'Same analysis but specifying an unstructured covariance matrix.';

**Proc** **Mixed**; class group subject time;

model dv = group time group\*time;

repeated time /subject = subject type = un hlm hlps rcorr; **run**;

Title 'Same analysis but specifying an autoregressive covariance matrix.';

**Proc** **Mixed**; class group subject time;

model dv = group time group\*time;

repeated time /subject = subject type = AR(**1**) rcorr; **run**;

|  |
| --- |
| **The Proc Mixed analysis with missing data and assumed compound symmetry.** |

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HOWELL |
| **Dependent Variable** | dv |
| **Covariance Structure** | Compound Symmetry |
| **Subject Effect** | subject |
| **Estimation Method** | REML |
| **Residual Variance Method** | Profile |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **group** | 2 | 1 2 |
| **subject** | 24 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 |
| **time** | 4 | 1 2 3 4 |

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 2 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 24 |
| **Max Obs Per Subject** | 4 |

| **Number of Observations** | |
| --- | --- |
| **Number of Observations Read** | 87 |
| **Number of Observations Used** | 87 |
| **Number of Observations Not Used** | 0 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 924.61183608 |  |
| **1** | 2 | 905.39818761 | 0.00000178 |
| **2** | 1 | 905.39750232 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

| **Estimated R Correlation Matrix for subject 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.4640 | 0.4640 | 0.4640 |
| **2** | 0.4640 | 1.0000 | 0.4640 | 0.4640 |
| **3** | 0.4640 | 0.4640 | 1.0000 | 0.4640 |
| **4** | 0.4640 | 0.4640 | 0.4640 | 1.0000 |

| **Covariance Parameter Estimates** | | |
| --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** |
| **CS** | **subject** | 2558.27 |
| **Residual** |  | 2954.66 |

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 905.4 |
| **AIC (smaller is better)** | 909.4 |
| **AICC (smaller is better)** | 909.6 |
| **BIC (smaller is better)** | 911.8 |

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 1 | 19.21 | <.0001 |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **group** | 1 | 22 | 16.53 | 0.0005 |
| **time** | 3 | 57 | 32.45 | <.0001 |
| **group\*time** | 3 | 57 | 6.09 | 0.0011 |

|  |
| --- |
| **Same analysis but specifying an unstructured covariance matrix.** |

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HOWELL |
| **Dependent Variable** | dv |
| **Covariance Structure** | Unstructured |
| **Subject Effect** | subject |
| **Estimation Method** | REML |
| **Residual Variance Method** | None |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 10 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 24 |
| **Max Obs Per Subject** | 4 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 924.61183608 |  |
| **1** | 4 | 885.22555000 | 0.00279889 |
| **2** | 1 | 883.95968197 | 0.00059787 |
| **3** | 1 | 883.70684919 | 0.00004177 |
| **4** | 1 | 883.69066243 | 0.00000026 |
| **5** | 1 | 883.69056547 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

| **Estimated R Correlation Matrix for subject 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.5858 | 0.5424 | -0.02740 |
| **2** | 0.5858 | 1.0000 | 0.8581 | 0.3896 |
| **3** | 0.5424 | 0.8581 | 1.0000 | 0.3971 |
| **4** | -0.02740 | 0.3896 | 0.3971 | 1.0000 |

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 883.7 |
| **AIC (smaller is better)** | 903.7 |
| **AICC (smaller is better)** | 906.9 |
| **BIC (smaller is better)** | 915.5 |

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 9 | 40.92 | <.0001 |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **group** | 1 | 22 | 17.95 | 0.0003 |
| **time** | 3 | 22 | 28.44 | <.0001 |
| **group\*time** | 3 | 22 | 6.80 | 0.0021 |

| **Type 3 Hotelling-Lawley-McKeon Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **time** | 3 | 20 | 25.85 | <.0001 |
| **group\*time** | 3 | 20 | 6.18 | 0.0038 |

| **Type 3 Hotelling-Lawley-Pillai-Samson Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **time** | 3 | 20 | 25.85 | <.0001 |
| **group\*time** | 3 | 20 | 6.18 | 0.0038 |

|  |
| --- |
| **Same analysis but specifying an autoregressive covariance matrix.** |

The Mixed Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HOWELL |
| **Dependent Variable** | dv |
| **Covariance Structure** | Autoregressive |
| **Subject Effect** | subject |
| **Estimation Method** | REML |
| **Residual Variance Method** | Profile |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| **Dimensions** | |
| --- | --- |
| **Covariance Parameters** | 2 |
| **Columns in X** | 15 |
| **Columns in Z** | 0 |
| **Subjects** | 24 |
| **Max Obs Per Subject** | 4 |

| **Iteration History** | | | |
| --- | --- | --- | --- |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 924.61183608 |  |
| **1** | 2 | 895.06992427 | 0.00000998 |
| **2** | 1 | 895.06615690 | 0.00000000 |

|  |
| --- |
| Convergence criteria met. |

| **Estimated R Correlation Matrix for subject 1** | | | | |
| --- | --- | --- | --- | --- |
| **Row** | **Col1** | **Col2** | **Col3** | **Col4** |
| **1** | 1.0000 | 0.6182 | 0.3822 | 0.2363 |
| **2** | 0.6182 | 1.0000 | 0.6182 | 0.3822 |
| **3** | 0.3822 | 0.6182 | 1.0000 | 0.6182 |
| **4** | 0.2363 | 0.3822 | 0.6182 | 1.0000 |

| **Covariance Parameter Estimates** | | |
| --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** |
| **AR(1)** | **subject** | 0.6182 |
| **Residual** |  | 5350.25 |

| **Fit Statistics** | |
| --- | --- |
| **-2 Res Log Likelihood** | 895.1 |
| **AIC (smaller is better)** | 899.1 |
| **AICC (smaller is better)** | 899.2 |
| **BIC (smaller is better)** | 901.4 |

* A smaller fit statistic indicates better fit between model and data. Howell suggests using these to help select which model to adopt. The AIC criterion had a value of 909.4 for the model assuming compound symmetry, 909.7 when no structure was imposed, and 899.1 with the autoregressive model.

| **Null Model Likelihood Ratio Test** | | |
| --- | --- | --- |
| **DF** | **Chi-Square** | **Pr > ChiSq** |
| 1 | 29.55 | <.0001 |

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **group** | 1 | 22 | 17.32 | 0.0004 |
| **time** | 3 | 57 | 30.82 | <.0001 |
| **group\*time** | 3 | 57 | 7.72 | 0.0002 |

* [Zip File with code and data](http://core.ecu.edu/psyc/wuenschk/MV/RM-ANOVA/Mixed-Proc_ANOVA-1W1B.zip)

[Karl L. Wuensch](http://core.ecu.edu/psyc/WuenschK/KLW.htm), November, 2019. [Fair Use of this Document](http://core.ecu.edu/psyc/WuenschK/Fair-Use.htm)