

## 11.2: Correlated Residuals

### Note

The first part of the section uses a hypothetical data set to illustrate the origin of the covariance structure, by capturing the residuals for each time point and looking at the simple correlations for pairs of time points. *Therefore, the software code used for this purpose is NOT what we would ordinarily use in conducting a repeated measures analysis* as generating the residuals of a fitted model and their variances and covariances is automatically done by software. The variances and the covariances of the residuals will be outputted as the diagonals and the off-diagonals of the variance-covariance (R block) matrix in SAS or R. Minitab currently does not accommodate various covariance structures, opting instead to treat repeated measures as "split-plot in time" (which assumes compound symmetry).

If we look at the ANOVA mixed model in general terms, we have:

$$\text{Model: response} = \text{fixed effects} + \text{random effects} + \text{errors} \quad (11.2.1)$$

In the case of repeated measures with measures taken at  $p$  number of time points, the covariance structure of the errors can be expressed as a matrix. The diagonals of this matrix are the error variances at each time point. The off-diagonals are the covariances between successive time points. In general, the variance-covariance matrix can be expressed as follows:

$$\sum_i = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \quad (11.2.2)$$

The structure shown above does not assume any specific properties of the variances and covariances and is called an **unstructured covariance structure**. Note that there are  $p$  variances and  $\frac{p(p-1)}{2}$  covariances that adds to  $\frac{p(p+1)}{2}$  unknown quantities which define this matrix. So, even for a small number of time points, a substantial number of parameters will have to be estimated. Therefore, in practice, specific structures are imposed to reduce the number of distinct parameters that need to be estimated, which will be discussed in Section 11.3.

To understand the correlation structure of errors, let us use SAS to generate the variance-covariance matrix of the errors for a repeated measures model using hypothetical data stored in [Repeated Measures Example Data](#). The data consists of a single treatment with 3 levels. Subjects are assigned a treatment level at random (CRD) and then are measured at  $p = 3$  time points. The SAS code which is given below fits a factorial model and generates the errors along with the correlations among responses taken at three time points.

```
data rmanova;
  input trt $ time subject resp;
  datalines;
  A 1 1 10
  A 1 2 12
  A 1 3 13
  A 2 1 16
  A 2 2 19
  A 2 3 20
  A 3 1 25
  A 3 2 27
  A 3 3 28
  B 1 4 12
  B 1 5 11
  B 1 6 10
```

```

B 2 4 18
B 2 5 20
B 2 6 22
B 3 4 25
B 3 5 26
B 3 6 27
C 1 7 10
C 1 8 12
C 1 9 13
C 2 7 22
C 2 8 23
C 2 9 22
C 3 7 31
C 3 8 34
C 3 9 33
;

```

We can run a simple model and obtain the residuals.

```

/* 2-factor factorial for trt and time - saving residuals */
proc mixed data=rmanova method=type3;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr outpm=outmixed;
  title 'Two_factor_factorial';
run; title;

```

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	2	18	14.52	F">0.0002
time	2	18	292.72	F"><.0001
trt*time	4	18	4.67	F">0.0092

```

/* re-organize the residuals to (unstacked data for correlation) */
data one;
  set outmixed;
  where time=1; time1=resid;
  keep time1;
run;
data two; set outmixed; where time=2; time2=resid; keep time2; run;
data three; set outmixed; where time=3; time3=resid; keep time3; run;

data corrcheck; merge one two three;

proc print data=corrcheck;
run;

```

```
proc corr data=corrcheck nosimple; var time1 time2 time3; run;
```

The residuals then are:

The Print Procedure			
Obs	time1	time2	time3
1	-1.66667	-2.33333	-1.66667
2	0.33333	0.66667	0.33333
3	1.33333	1.66667	1.33333
4	1.00000	-2.00000	-1.00000
5	0.00000	0.00000	0.00000
6	-1.00000	2.00000	1.00000
7	-1.66667	-0.33333	-1.66667
8	0.33333	0.66667	1.33333
9	1.33333	-0.33333	0.33333

The correlations of responses between time points are:

#### The CORR Procedure

3 Variables: time1 time2 time3

Pearson Correlation Coefficients, N = 9

Prob > |r| under H0: Rho=0

	time1	time2	time3
time1	1.00000	0.19026 0.6239	0.55882 0.1178
time2	0.19026 0.6239	1.00000	0.83239 0.0054
time3	0.55882 0.1178	0.83239 0.0054	1.00000

Notice that in the above code, the repeated nature of the data is not being utilized. The "*repeated*" statement in `proc mixed`, which is used in practice, accounts for this. As in the code given below, in the repeated statement, the option of `subject=` specifies what experimental (or observational) units the repeated measures are made on. The `type=` can be used to specify one of many types of structures for these correlations. Here we specified the unstructured covariance structure and obtained the same correlations that were generated earlier with simple statistics.

```
proc mixed data=rmanova ;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr solution ;
  repeated /subject=subject(trt) type=UN rcorr;
  title 'Repeated Measures';
run; title;
```

Finding the best covariance structure is much of the work in modeling repeated measures and is usually done by considering a subset of candidate structures. These include UN (Unstructured), CS (Compound Symmetry), AR(1) (Autoregressive lag 1) – if time intervals are evenly spaced, or SP(POW) (Spatial Power) – if time intervals are unequally spaced.

Choosing the best covariance structure is based on Fit Statistics (also known as information criteria). PROC MIXED in SAS automatically generates four of such Fit Statistics measures and for this example, they are:

Fit Statistics	
<b>-2 Res Log Likelihood</b>	63.0
<b>AIC (Smaller is Better)</b>	75.0
<b>AICC (Smaller is Better)</b>	82.6
<b>BIC (Smaller is Better)</b>	76.2

Smaller or more negative values indicate a better fit to the data. The process amounts to trying various candidate structures and then selecting the covariance structure producing the smallest or most negative values. The information criteria listed above are usually similar in value, but for small sample sizes, the AICC criterion is recommended. The topic of covariance structures for a general setting is discussed in the next section.

## ? Using R

- Load the Repeated Measures Example Data.
- Obtain the ANOVA table.
- Obtain the correlations of responses between time points.
- Obtain the results for the split-plot in time approach.
- Run the analysis as a repeated-measures ANOVA by using different covariance structures.

### Steps in R

1. Load the Repeated Measures Example data and obtain the ANOVA by using the following commands:

```
setwd("~/path-to-folder/")
repeated_measures_example_data <- read.table("repeated_measures_example_data.txt")
attach(repeated_measures_example_data)
rmanova<-aov(resp ~ trt + factor(time) + trt*factor(time), repeated_measures_example_data)
anova(rmanova)
#Analysis of Variance Table
#Response: resp
#
#          Df    Sum Sq   Mean Sq    F value    Pr(>F)
#trt         2     64.52     32.26    14.5167    0.0001761 ***
#factor(time)  2  1300.96    650.48   292.7167    1.87e-14 ***
#trt:factor(time)  4    41.48     10.37     4.6667    0.0092424 **
#Residuals   18    40.00      2.22
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2. Obtain the correlations of responses between time points by using the following commands:

```
time_1<-c(rmanova$residuals[1:3],rmanova$residuals[10:12],rmanova$residuals[19:21])
time_2<-c(rmanova$residuals[4:6],rmanova$residuals[13:15],rmanova$residuals[22:24])
time_3<-c(rmanova$residuals[7:9],rmanova$residuals[16:18],rmanova$residuals[25:27])
residuals<-cbind(time_1,time_2,time_3)
```

```
rownames(residuals)<-NULL
#residuals
#           time_1           time_2           time_3
#[1,] -1.666667e+00 -2.333333e+00 -1.666667e+00
#[2,]  3.333333e-01  6.666667e-01  3.333333e-01
#[3,]  1.333333e+00  1.666667e+00  1.333333e+00
#[4,]  1.000000e+00 -2.000000e+00 -1.000000e+00
#[5,] -3.885781e-16  9.436896e-16 -7.216450e-16
#[6,] -1.000000e+00  2.000000e+00  1.000000e+00
#[7,] -1.666667e+00 -3.333333e-01 -3.333333e-01
#[8,]  3.333333e-01  6.666667e-01  6.666667e-01
#[9,]  1.333333e+00 -3.333333e-01 -3.333333e-01
#cor(residuals)
#
#           time_1           time_2           time_3
#time_1  1.0000000  0.1902606  0.3290726
#time_2  0.1902606  1.0000000  0.9756655
#time_3  0.3290726  0.9756655  1.0000000
```

This page titled [11.2: Correlated Residuals](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.