Repeated Measures Covariance Structure

Alex Lipka
<alipka@stat.purdue.edu>

Benjamin Tyner
<btyner@stat.purdue.edu>

November 7, 2004

Statistical Consulting Service
Purdue University
West Lafayette, IN, USA

1 Repeated Measures

Any measurement that can be repeated (either across time or across space) can be analyzed under this broad heading. Crowder and Hand[2] describe *repeated measures* as the situation in which measurements "are made of the same characteristic on the same observational unit but on more than one occasion." This is what is meant by the term *longitudinal data*. The scope of repeated measures can be expanded to include *clustered data* as well; that is, measurements on members of a cluster that are related in some way. For longitudinal data, a common occurance is for measurements on the same observational unit to be correlated; for cluster data, the same phenomenon can occur, with clusters playing the role of observational units, and repetition of measurement occuring within a cluster. In both cases, the usual model assumption of independent errors may be violated, so a model that can incorporate this lack of independence is needed.

2 Covariance Structure

We assume N observational units (individuals or clusters) and n_i observations of the response for the i^{th} unit, $i = 1, \ldots, N$ (if there are no missing values, then $n_i = n$). The observations for the i^{th} unit are coded in the vector $\boldsymbol{y}_i$, which has length n_i. The design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ consist of q columns and n_i rows. The entries of $\boldsymbol{X}_i$ are the values taken by the q continuous explanatory variables and/or design indicators of fixed effects across the n_i measurements of the i^{th} unit; if any between-subjects random effects exist, their design indicators appear in $\boldsymbol{Z}_i$. For the model

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i$$

$$\begin{bmatrix} \boldsymbol{\gamma}_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{G}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R}_i \end{bmatrix}\right)$$

the term Covarianace Structure is used to describe how the matrices $\boldsymbol{G}_i$ and $\boldsymbol{R}_i$ are constrained in the (Normal case of the) general linear mixed model:

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{\text{ind}}{\sim} \mathcal{N}_{n_i}\left(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{V}_i\right)$$

where $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{Z}_i^{\text{T}} + \boldsymbol{R}_i$ is called the *Variance-Covariance Matrix* of the i^{th} unit. It is important to note that *the decomposition of* $\mathbf{V}_i$ *into the* $\mathbf{G}_i$ *term and the* $\boldsymbol{R}_i$ *term is not necessarily unique*; that is, there may be two or more structures for $\boldsymbol{G}_i$ and $\boldsymbol{R}_i$ that yield the same $\boldsymbol{V}_i$. An example of this phenomenon will be given later. Typically, it is assumed that all of the $\boldsymbol{V}_i$s take the same form[3]. The examples that follow are forms of $\boldsymbol{R}_i$, but can be used for $\boldsymbol{G}_i$ as well:

2.1 Unstructured

This is the most general form:

$$\boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n_i} \\ \rho_{21} & 1 & & \rho_{2n_i} \\ \vdots & & \ddots & \vdots \\ \rho_{n_i1} & \rho_{n_i2} & \cdots & 1 \end{bmatrix}$$

In this case the varaince-covariance matrix contains $n_i\left(n_i - 1\right)/2 + 1$ nuisance parameters to be estimated, so in practice, estimation of this structure may only convergence for $N \gg n_i$. Likewise the statistical power under this structure is reduced since the only "constraint" on $\boldsymbol{R}_i$ is that it be symmetric.

2.2 ARMA(1,1)

The more structured forms of $\boldsymbol{V}_i$ involve assuming that some or all of the ρ_{jk}s are a function of the "distance" between observations j and k. If measurements are repeated across time, then the time variable will be prominent in this distance function. If measurements are repeated across location, then the distance function will involve some spatial metric reflecting the experimental design's geometry. One of the more common structures is the *first order autoregressive, first order moving-average model*, abbreviated ARMA(1,1):

$$\boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & \gamma\rho^0 & \cdots & \gamma\rho^{n_i-2} \\ \gamma\rho^0 & 1 & & \gamma\rho^{n_i-3} \\ \vdots & & \ddots & \vdots \\ \gamma\rho^{n_i-2} & \gamma\rho^{n_i-3} & \cdots & 1 \end{bmatrix}$$

Notice that now there are only three parameters to be estimated, σ^2, γ, and ρ. A special case of ARMA(1,1) occurs when $\gamma = \rho$; this is called ARMA(1,0) or just simply AR(1), and only two covariance parameters σ^2 and ρ are estimated.

2.3 Equicorrelation

Another model assumes all repeated measurements are equally correlated:

$$\boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

This structure (also called *spherical* or *exchangeable*) might be applicable to cluster data, where ρ is called the *intra-class correlation coefficient* between two members of the same cluster and is "a relative measure of the within-cluster similarity."[3] A special case of equicorrelation, called *compound symmetry*, arises by enforcing $\rho = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$ for some σ_a^2 and σ_e^2. In that case, if $\sigma^2 = \sigma_a^2 + \sigma_e^2$, then

$$\boldsymbol{R}_i = \begin{bmatrix} \sigma_a^2+\sigma_e^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2+\sigma_e^2 & & \sigma_a^2 \\ \vdots & & \ddots & \vdots \\ \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2+\sigma_e^2 \end{bmatrix}$$

2.4 Uncorrelated

Finally, we compare to the fixed effects case where independence is assumed over repeated measurements:

$$\boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

This helps to show (when $\boldsymbol{G}_i = \boldsymbol{0}$) that the standard linear model can be thought of as a special case of repeated measures.

3 Software Packages

3.1 Repeated Measures in SAS

Both `PROC GLM` and `PROC MIXED` provide a `REPEATED` statement by which one or more repeated measures can be specified. However, there are many nontrivial differences between these two procedures. Some of the more important differences are:

- `PROC GLM` has as a model assumption that $\boldsymbol{R}_i$ is either unstructured or of "Type H" structure[5], whereas `PROC MIXED` allows a choice between over 30 different structures for $\boldsymbol{R}_i$ through the `TYPE=` option of the `REPEATED` statement.

- `PROC GLM` ignores observational units with missing observations, whereas `PROC MIXED` includes them. (Though note that they must be missing at random for the estimators to remain unbiased.)

- `PROC GLM` assumes covariates are constant within observational units, whereas in `PROC MIXED` they are allowed to vary.

See[6] for a longer treatment of the differences. Note that if `REPEATED` is omitted, SAS uses the uncorrelated $\boldsymbol{R}_i$ from section (2.4).

3.1.1 Unstructued

Seven subjects had their response times measured when a light was flashed into each eye through lenses of powers 6/6, 6/18, 6/36, and 6/60 (a lens of power a/b means that the eye will perceive as being at a feet an object actually positioned at b feet). Measurements were made in milliseconds, and the question of interest was whether response time varied with lens strength.[4] Note that there are not enough subjects to fit an unstructured model for both left and right eyes.

```
Data Vision; /* From Everitt 1996:103 */
 Input Subject @;
  Do Eye="Left ","Right";
   Do LensStrength="6/06","6/18","6/36","6/60";
    Input ResponseTime @;
Output;
   End;
  End;
DataLines;
1     116    119    116    124    120    117    114    122
2     110    110    114    115    106    112    110    110
3     117    118    120    120    120    120    120    124
4     112    116    115    113    115    116    116    119
5     113    114    114    118    114    117    116    112
6     114    115     94    116    100     99     94     97
7     110    110    105    118    105    105    115    115
;
run;

Proc Mixed Data=Vision;
```

```
 Where Eye="Left ";
 Class Subject LensStrength;
 Model ResponseTime = LensStrength;
 Repeated LensStrength / Subject=Subject Type=UN R RCorr;
 LSMeans LensStrength / PDiff;
Run;
```

3.1.2 AR(1)

The milk production of four cows is followed over four lactation periods and four diets; assume that diet does not have any residual effect.[7]

```
data new;
 input cow period trt resp @@;
cards;
 1 1 1 38 1 2 2 32 1 3 3 35 1 4 4 33
 2 1 2 39 2 2 3 37 2 3 4 36 2 4 1 30
 3 1 3 45 3 2 4 38 3 3 1 37 3 4 2 35
 4 1 4 41 4 2 1 30 4 3 2 32 4 4 3 33
;
run;

proc mixed data = new;
class cow trt period;
model resp = trt period;
random cow;
repeated period / type = ar(1) subject = cow;
run;
```

3.1.3 Compound Symmetry

Over the course of a month, three different methods of throwing a baseball were taught to 21 subjects with seven subjects per method. The throwing speed for each subject was recorded at two and four weeks and adjusted for their initial throwing speed.[1] Note that subjects are nested within method:

```
data new;
input meth subj time1 time2;
cards;
1    1 25.4 30.6
1    2 27.4 29.3
1    3 25.5 30.0
1    4 25.8 29.7
1    5 26.2 31.3
1    6 24.6 26.6
1    7 25.6 28.0
2    1 27.6 27.1
2    2 24.7 29.0
2    3 26.3 27.3
```

```
2    4 25.0 29.7
2    5 25.7 29.5
2    6 28.5 29.7
2    7 22.9 27.2
3    1 22.8 25.1
3    2 24.2 24.0
3    3 25.3 25.2
3    4 25.4 24.7
3    5 24.5 26.2
3    6 25.6 26.9
3    7 25.6 24.8
;
run;

data new1;
 set new;
 resp=time1; time=1; output;
 resp=time2; time=2; output;
run ;

proc mixed data=new1;
 class meth subj time;
 model resp= meth|time;
 repeated / type = cs sub = subj;
run;
```

3.2 Repeated Measures in SPSS

(To be completed)

References

[1] B.A. Craig. Stat 514: Experimental design class notes, topic 22. 2004.

[2] M.J. Crowder and D.J. Hand. *Analysis of Repeated Measures*. Chapman & Hall, 1990.

[3] Annette J. Dobson. *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC, 2002.

[4] B.S. Everitt. *Making Sense of Statistics in Psychology: A Second-Level Course*. Oxford University Press, Oxford, 1996.

[5] SAS Institute Inc., Cary, NC. *SAS OnlineDoc®, Version 8*, 1999.

[6] R.D. Wolfinger and M. Chang. Comparing the SAS GLM and MIXED procedures for repeated measures. Technical report, SAS Institute Inc., Cary, NC, 1998.

[7] Y. Zhu. Stat 514: Experimental design class notes, latin square and related designs. 2004.