Covariance Structure Approach to Within-cases

Using SAS `proc mixed`

Weaknesses of the classical mixed model approach to repeated measures

- Requires a balanced design for the “F” statistics to have F distributions under H_0.
- If data are collected over time, there is trouble if the observations are unequally spaced.
- If different numbers of observations are collected for different individuals, there is trouble.
- Can’t have covariates.
- The model may not reflect reality for some data sets.

The classical mixed model

- The “random effect” for subjects is a little piece of random error, characteristic of an individual.
- If that’s the only reason for correlation between measurements from the same case, the performance of the tests is excellent.
- If there are other sources of covariance among the repeated measures (like learning, or fatigue, or memory of past performance), there is too much chance of wrongly rejecting the null hypothesis (Type I error).
- The Greenhouse-Giesser correction compensates for the problem by adjusting the degrees of freedom.
- Often it’s an over-correction, resulting in diminished power.

Toward a solution

- Why are data from the same case correlated in the classical approach?
- Because each case makes its own contribution -- add a (random) quantity that is different for each case.
- So variances of measurements are all equal.
- And correlations are all equal.
- Classical mixed model for within cases implies *compound symmetry*.

Compound Symmetry

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

`Proc mixed` directly incorporates more general covariance structures.

Advantages of the covariance structure approach

- Straightforward: It's familiar univariate regression.
- Just MSE is different, because of correlated observations.
- Nicer treatment of missing data (valid if missing at random).
- Can have time-varying covariates.
- Flexible modeling of non-independence within cases.
- Can accommodate more factor levels than cases (with assumptions).

Usual covariance matrix of $Y_1, \ldots, Y_n$

$$\begin{bmatrix} \sigma^2 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \sigma^2 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \sigma^2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & 0 & \ldots & \sigma^2 & 0 \\ 0 & 0 & 0 & \ldots & 0 & \sigma^2 \end{bmatrix}$$

In the covariance structure approach

- There are *n* "subjects."
- There are *k* ("repeated") measurements per subject
- There are *nk* cases: *n* blocks of *k* rows
- Data are multivariate normal (dimension *nk*)
- Familiar regression model for $E(Y_i|\mathbf{X})$
- Special structure for the variance-covariance matrix: not just a diagonal matrix with σ^2 on the main diagonal.

Structure of the variance-covariance matrix of $Y_1, \ldots, Y_n$

- Covariance matrix of the data has a **block diagonal** structure: *nxn* matrix of little k*xk* variance-covariance matrices (partitioned matrix).
- Off diagonal matrices are all zeros -- no correlation between data from different cases
- Matrices on the main diagonal are all the same (equal variance assumption)

Block Diagonal Covariance Matrix

$$\left[\begin{array}{c|c|c|c|c} \Sigma & 0 & \ldots & 0 & 0 \\ \hline 0 & \Sigma & \ldots & 0 & 0 \\ \hline \vdots & \vdots & \ddots & \vdots & 0 \\ \hline 0 & 0 & \ldots & \Sigma & 0 \\ \hline 0 & 0 & \ldots & 0 & \Sigma \end{array}\right]$$

Σ is the matrix of variances and covariances of the data from a single subject.

$\boldsymbol{\Sigma}$ may have different *structures*

- May be unknown

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_4^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ & & \sigma_3^2 & \sigma_{3,4} \\ & & & \sigma_4^2 \end{bmatrix}$$

- May be something else

Available covariance structures include

- Unknown: type=un
- Compound symmetry: type=cs
- Variance components: type=vc
- First-order autoregressive: type=ar(1)
- Spatial autocorrelation: covariance is a function of Euclidian distance
- Factor analysis
- Many others

Compound Symmetry

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

Advantage: Fewer parameters to estimate

Why not always assume covariance structure unknown?

- No reason why not, if you have enough data.
- When number of unknown parameters is large relative to sample size, variances of estimators are large, so confidence intervals wide, tests weak.
- In some studies, there can be more treatment conditions than cases, and unique estimates of parameters don’t even exist.
- There is always a tradeoff between assumptions and amount of data.

First-order autoregressive time series

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

- Usually much bigger matrix
- Could have a handful of cases measured at hundreds of time points
- Or even just one "case," say a company

Eating Norm Study

- Two free meals at the psych lab (on different days)
- One with another student, one alone
- But it’s not really another student. It’s a “confederate.”
- Confederate either eats a lot or a little.
- Dine with the confederate first, or second.
- Response variable is how much you eat. They weigh it.
- Covariates: How long since you ate, and how hungry you are. (Self Report)
- Classical approach just can’t handle it.
- Neither can the multivariate approach.

Variables

- Amount subject eats: DV
- Amount confederate eats (between)
- Eat alone or with confederate (within)
- Eat with confederate first, or second (between)
- Reported time since ate (covariate)
- Reported hunger (covariate)

- Notice these are **time-varying covariates**

Copyright Information