Selecting The Covariance Structure

Covariance Structures

The major difference between a univariate regression model for independent observations and a multivariate model for repeated measures is that the assessments for each individual are assumed to be correlated over time, $Var[Yi] = \Sigma i$. Examples of covariance structures include unstructured, Toplitz, compound symmetry (exchangeable) and autoregressive.

Variance Components

The VC structure is the standard variance components and is the default.

$$\begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$$

Autoregressive(1)

The AR(1) structure has homogeneous variances but correlations decline exponentially with distance. In our case this means that the variability in a measurement, say white blood cell count, is constant regardless of when you measure it. It also means that two measurements that are right next to each other in time are going to be pretty correlated (depending on the value of ρ), but that as measurements get farther and farther apart they are less correlated.

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Compound symmetry

The CS structure is the well-known compound symmetry structure required for split-plot designs "in the old days". As can be seen in the table, the variances are homogeneous. There is a correlation between two separate measurements, but it is assumed that the correlation is constant regardless of how far apart the measurements are.

Unstructured

$$\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{bmatrix}$$

The UN structure is the most "liberal" of all allowing every term to be different. It requires fitting the most parameters of any structure, t(t+1)/2. The unstructured covariance is the least restrictive and generally the best choice when the number of repeated measures is small. When the number of repeated measures increases, the number of parameters in the unstructured covariance matrix increases dramatically.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}$$

Toeplitz

The TOEP structure is similar to the AR(1) in that all measurements next to each other have the same correlation, measurements two apart have the same correlation different from the first, measurements three apart have the same correlation different from the first two, etc. However, the correlations do not necessarily have the same pattern as in the AR(1). Technically, the AR(1) is a special case of the Toeplitz.

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

First order Autoregressive Moving Average (ARMA(1,1))

ARMA(1,1) provide a more flexible structures that more closely fit the correlation typical of HRQOL (Health-related quality of life) measures.

$$\begin{bmatrix} \sigma^2 & \sigma^2\lambda & \sigma^2\lambda\rho & \sigma^2\lambda\rho^2 \\ \sigma^2\lambda & \sigma^2 & \sigma^2\lambda & \sigma^2\lambda\rho \\ \sigma^2\lambda\rho & \sigma^2\lambda & \sigma^2 & \sigma^2\lambda \\ \sigma^2\lambda\rho^2 & \sigma^2\lambda\rho & \sigma^2\lambda & \sigma^2 \end{bmatrix}$$

Heterogeneous versions of the above

The heterogeneous versions of the covariance structures above are a simple extension. That is the variances, along the diagonal of the matrix, do not have to be the same. Note that this adds more parameters to be estimated, one for every measurement.

Recommendations: If the correlation between measures is roughly equal regardless of how far apart in time the observations are taken, then compound symmetry is a possible candidate. When the correlation of the residual errors is likely to be strongest for observations that are close in time and weakest for observations that are the furthest apart. The Toplitz, the autoregressive moving average (ARMA(1,1)) and the autoregressive structure (AR(1)) allow observations that are further apart to be less strongly correlated. The autoregressive is the most restrictive of these structures.

Objectives

• To identify possible choices for the covariance associated with **Repeated Measures** models.
• To produce a parsimonious model if desired.
• To construct simple and complex tests of hypotheses.

Strategies

Once you have the random effects determined, then you can move on to selecting the covariance structure. There are a variety of considerations when selecting the covariance structure. They include the number of parameters, the interpretation of the structure, diagnostic results, and effects on fixed effects.

By Parsimony (Economically)

If the data suffices, one could always fit the unstructured covariance structure and go with it. However, just as in traditional regression we want to have as few parameters in the model as possible. The more data you have the more parameters you can fit, but they do not always add to our knowledge and often take away. The more complex the model the more specific to the data it will be and the less generalizable. Our belief that Nature follows simple, elegant rules leads us to look for the simpler model when possible. From the fixed effects perspective, selecting a structure that is too simple increases the fixed effects Type I error rate, and selecting a structure that is too complex sacrifices power and efficiency.

By Meaning

Using your understanding of the design and treatment structures and the meaning of the covariance structures will usually give you a few candidate structures to work with. Therefore, even though choosing the candidate structures before the data is collected, i.e. as part of the SAP (Business application technology), may be required in a clinical trial, it is definitely the best method, since it avoids allowing chance to drive the covariance estimates.

By IC (Information Criteria)

Specifying the IC option on the PROC (abbreviation that refers to weapon, item or ability activating with the) statement gives three default Information Criteria.

- AIC = Akaike's Information Criteria,
- AICC = AIC Corrected, and
- BIC = Bayesian Information Criteria.

These statistics are functions of the log likelihood and can be compared across models, if you keep the fixed effects part of the model constant. SAS provides them in the smaller is better format. One

could fit models with competing covariance structures and compare the IC. They should give you comparable results and you can use the democratic rule to determine the best model. If they give conflicting results, then the simpler model is probably better.

Graphically

For us visual learners, there are a few graphical techniques that have been developed. Littell recommends to fit the data with an unstructured covariance matrix asking for the residual correlations or covariances. Plotting the covariances separately for each starting time can provide diagnostic information.

****