Solutions to Written Examination for Linear Mixed Models (MSA650 and MVE210)

Teacher: Ziad Taib, +46 707 655471 and Malin Östensson (+46 31 772 53 16)
Date and place: 2009-05-26 in V
Rules: This is a closed book exam. No material is allowed other than a simple pocket calculator.
Scores: The written exam is worth 24 scores (80%) while the computer assignments are worth 6 scores (20%). Together these two parts add up to 30 scores (100%). There are three possible overall grades: Excellent (at least 26 scores), pass (at least 16 cores) and do not pass (less than 16 cores).

---

1. See the book

2. The diagonal elements of the covariance matrix, which estimate overall variance in the population at each time point, appear to increase over time, suggesting that the true population variances may not be the same at each time point. This might suggest a heterogeneous model is appropriate, although one might argue that the increase is not profound enough to abandon a homogeneous assumption.
The sample correlation matrix suggests that observations one time interval (1.5 hours { these times are equally-spaced) are positively correlated and with correlation of similar magnitude (roughly 0.5), but observations 2 or more time intervals apart show negligible correlation, with all the estimates · 0:10 in absolute value. The correlations one time interval apart are not exactly the same, as this is an estimate, but they are in a similar “ballpark" suggesting that maybe the true population correlations could be the same. Likewise, the off-diagonal elements are very close to zero with the possible exception of 0.10; again, as this is a sample estimate, if the true correlation was zero, such an estimate might still be obtained.
These correlations are much smaller than the 1-time-interval correlations. These observations are consistent with a heterogeneous one-dependent covariance structure. AR(1) is also a possibility.

Does a particular source of variation/correlation appear to be dominant?" If so, identify the source and say why you think this is the case. If not, explain why you do not think so. Here, correlation drops substantially when observations are more than one time interval (1.5 hours) apart. This is characteristic of the tendency for deviations due to within-unit fluctuations" to become less alike the farther apart in time they are. This suggests that, in terms of contribution to the overall pattern of correlation, the within-unit source of correlation is dominant.

3.

Let $\boldsymbol{Y}_i$ be the vector of *observed* responses for a subject who missed his scheduled visits at months 2, 4, and 10.

(a) Write down the covariance matrix for $\boldsymbol{Y}_i$ if it is assumed that the covariance structure for the vector of *intended* responses follows a homogeneous AR(1) model.

The observations are equally-spaced, with a time interval of 2 months, so this model is reasonable. The *actual* times for this subject are (0,6,8,12), which correspond to intended times indexed by (1,4,5,7). So, for example, 0 and 6 are 3 time intervals apart, 8 and 12 are 2 intervals apart, and so on. With the homogeneity, assuming common variance σ^2 at all time points, we thus have:

$$\sigma^2 \begin{pmatrix} 1 & \rho^3 & \rho^4 & \rho^6 \\ & 1 & \rho^1 & \rho^3 \\ & & 1 & \rho^2 \\ & & & 1 \end{pmatrix},$$

where ρ is the correlation parameter for the AR(1) structure.

(b) Write down the covariance matrix for $\boldsymbol{Y}_i$ if it is assumed that the covariance structure for the vector of *intended* responses follows a heterogeneous compound symmetry model.

Here, variance changes with time. There are 7 *intended* time points, so let the (unequal) variances at times 0, 2, 4, 6, 8, 10, 12 be $\sigma_1^2, \sigma_2^2, \ldots, \sigma_7^2$. As above, we have only seen the observations at times indexed by (1,4,5,7). Thus, using this "intended" indexing, the matrix is

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_4 & \rho\sigma_1\sigma_5 & \rho\sigma_1\sigma_7 \\ & \sigma_4^2 & \rho\sigma_4\sigma_5 & \rho\sigma_4\sigma_7 \\ & & \sigma_5^2 & \rho\sigma_5\sigma_7 \\ & & & \sigma_7^2 \end{pmatrix},$$

where ρ is the assumed constant correlation for the compound symmetry correlation structure.

4.

$m = 15$

guinea pigs were randomly assigned to 3 groups, 5 pigs per group. At the beginning of the study (baseline, week 0) all pigs were given the same growth-inhibiting substance and were then treated identically until the end of week 4 (week 4). At this point, they were started on one of three vitamin E supplement doses, depending on the group to which they had been randomized (zero, low, or high dose). For each pig, body weight (g) was recorded at weeks 1, 3, 4, 5, 6, and 7.

For group 2, the low dose group, the model says that mean body weight *before* week 4 is $\beta_{02} + \beta_{12} t_{ij}$. The model also says that mean body weight *after* week 4 is

$$\beta_{02} + \beta_{12} t_{ij} + \beta_{22}(t_{ij} - 4) = \beta_{02} + \beta_{12}(4) + (\beta_{12} + \beta_{22})(t_{ij} - 4).$$

(This is similar for the other groups.) So if we take week 4 as the "origin," for the phase *after* week 4, the rate of change for group 2 is

$$\beta_{12} + \beta_{22}.$$

(b) Because the pigs in all three groups were treated *identically* until the end of week 4, the investigators wondered whether or not a simpler model that reflects this fact could be adopted. Collect all the parameters that describe the mean body weight trajectories in model (2) on the previous page in a parameter vector $\boldsymbol{\beta}$, and give the form of $\boldsymbol{\beta}$. Then write down the matrix $\boldsymbol{L}$ corresponding to the null hypothesis of the form $H_0 : \boldsymbol{L\beta} = \boldsymbol{0}$ addressing this issue.

Let

$$\boldsymbol{\beta} = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23})'.$$

As we discussed in part (a), the model in group $k = 1, 2, 3$ prior to week 4, while pigs were all treated identically, is

$$\beta_{0k} + \beta_{1k} t_{ij}.$$

If the pigs were treated identically until the end of week 4, we would expect the means across the three groups to be the same at any $t_{ij} \leq 4$. The only way this could be the case is if all the β_{0k} are the same and all the β_{1k} are the same. The $\boldsymbol{L}$ matrix is

$$\boldsymbol{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

(c) The investigators' key question was whether or not the pattern of change in mean body weight *after* the groups were put on their assigned vitamin E doses was the same in all groups. *Assuming that your null hypothesis in (b) is true*, write down a new model that incorporates this. Collect all the parameters that describe the mean body weight trajectories in your new model into a parameter vector $\boldsymbol{\beta}$, and give the form of $\boldsymbol{\beta}$. Then write down the matrix $\boldsymbol{L}$ corresponding to the null hypothesis of the form $H_0 : \boldsymbol{L\beta} = \boldsymbol{0}$ addressing this issue.

Under the hypothesis in (b), the model becomes

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_{21}(t_{ij} - 4)_+ + \epsilon_{ij}, \quad i \text{ from zero dose group} \\ &= \beta_0 + \beta_1 t_{ij} + \beta_{22}(t_{ij} - 4)_+ + \epsilon_{ij}, \quad i \text{ from low dose group} \\ &= \beta_0 + \beta_1 t_{ij} + \beta_{23}(t_{ij} - 4)_+ + \epsilon_{ij}, \quad i \text{ from high dose group,} \end{aligned}$$

so that the mean trajectories before week 4 are identical. We have

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \beta_{23})'.$$

The slopes after week 4 are thus $\beta_1 + \beta_{2k}$ for each group $k = 1, 2, 3$, and the question is whether these slopes are identical. This will be the case if $\beta_{21} = \beta_{22} = \beta_{23}$, which gives

$$\boldsymbol{L} = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix}.$$

In fact, if this null hypothesis is true, the profiles are all identical (with possibly different slopes before and after 4 weeks).

5. See the book

6. Recall from the notes that:

1. $E(Y_i) = \mu_i$
2. $\eta(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$ with $\eta(.)$ the **link function**
3. Var$(Y_i) = \phi v(\mu_i)$, where
   - $v(.)$ is a known **variance function**
   - ϕ is a scale (overdispersion) parameter
   - exponential family p.d.f.

     $$f(y|\theta_i, \phi) = \exp\left\{\phi^{-1}[y\theta_i - \psi(\theta_i)] + c(y, \phi)\right\}$$

     with θ_i the natural parameter and $\psi(.)$ a function satisfying
     - $\triangleright\ \mu_i = \psi'(\theta_i)$
     - $\triangleright\ v(\mu_i) = \psi''(\theta_i)$

If we assume the observations YB_1, Y_2, . . ., Y_N have a Poisson distribution then it is not difficult to see that this is indeed an exponential family and that the link function is $\psi(\mu) = \ln \mu$, so we have the Poisson regression model:

$$\ln\ \mu(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

to which we could add some random coefficients.