ST 790, Homework 5 Solutions
Spring 2018

1. If all the $\lambda_j(\mathcal{H}_{i,j-1})$ are correctly specified and MAR holds, then

$$\overline{\pi}_j(\mathcal{H}_{i,j-1})\lambda_{j+1}(\mathcal{H}_{ij}) = \text{pr}(D = j+1|\widetilde{\mathcal{Z}}_i, \boldsymbol{x}_i) = \text{pr}(D = j+1|\mathcal{H}_{ij})$$

Then consider the ith summand, which is equal to

$$\sum_{j=1}^{n} w_{ij}\boldsymbol{X}_i^{(j)T}(\beta)\{\boldsymbol{V}_i^{(j)}(\beta,\xi,\boldsymbol{x}_i)\}^{-1}\begin{pmatrix} Y_{i1} - f_1(\boldsymbol{a}_i,\beta) \\ \vdots \\ Y_{ij} - f_j(\boldsymbol{a}_i,\beta)\end{pmatrix}.$$

For the kth element, $k = 1,\ldots,j$,

$$\begin{aligned} &E[w_{ij}\boldsymbol{X}_i^{(j)T}(\beta)\{\boldsymbol{V}_i^{(j)}(\beta,\xi,\boldsymbol{x}_i)\}^{-1}\{Y_{ik} - f_1(\boldsymbol{a}_i,\beta)\}|\boldsymbol{x}_i] \\ &\quad = E\Big(E[w_{ij}\boldsymbol{X}_i^{(j)T}(\beta)\{\boldsymbol{V}_i^{(j)}(\beta,\xi,\boldsymbol{x}_i)\}^{-1}\{Y_{ik} - f_1(\boldsymbol{a}_i,\beta)\}|\widetilde{\mathcal{Z}}_i,\boldsymbol{x}_i]|\boldsymbol{x}_i\Big).\end{aligned}$$

Because $\widetilde{\mathcal{Z}}_i$ contains Y_{ik} for $k = 1,\ldots,j$ when $D = j+1$, the inner expectation becomes

$$\boldsymbol{X}_i^{(j)T}(\beta)\{\boldsymbol{V}_i^{(j)}(\beta,\xi,\boldsymbol{x}_i)\}^{-1}\{Y_{ik} - f_1(\boldsymbol{a}_i,\beta)\}\frac{E\{I(D=j+1)|\widetilde{\mathcal{Z}}_i,\boldsymbol{x}_i\}}{\text{pr}(D=j+1|\mathcal{H}_{ij})}.$$

But by MAR, $E\{I(D=j+1)|\widetilde{\mathcal{Z}}_i,\boldsymbol{x}_i\} = \text{pr}(D=j+1|\mathcal{H}_{ij})$, so that the inner expectation is

$$\boldsymbol{X}_i^{(j)T}(\beta)\{\boldsymbol{V}_i^{(j)}(\beta,\xi,\boldsymbol{x}_i)\}^{-1}\{Y_{ik} - f_1(\boldsymbol{a}_i,\beta)\}$$

Taking the outer expectation with respect to $\boldsymbol{x}_i$, using

$$E[\{Y_{ik} - f_1(\boldsymbol{a}_i,\beta)\}|\boldsymbol{x}_i] = 0,$$

and applying this argument to each $k = 1,\ldots,j$ for $j = 1,\ldots,n$ yields the result.

2. Write $\boldsymbol{R}_i$, $\boldsymbol{Z}_i$, etc for short here, so that

$$\widehat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{R}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{f}_i).$$

There are $\widehat{\boldsymbol{b}}_i$ in $\boldsymbol{Z}_i$ and $\boldsymbol{f}_i$, but this doesn't matter to the following argument, as all we are trying to do is reexpress (9.85) as (9.87).

There are two things to show. We can simplify (9.85) to

$$\begin{aligned} p(\boldsymbol{Y}_i|\boldsymbol{x}_i;\beta,\gamma,\boldsymbol{D}) &\approx (2\pi)^{-n_i/2}|\boldsymbol{R}_i|^{-1/2}|\boldsymbol{D}|^{-1/2}|\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i|^{-1/2} \\ &\quad \times \exp\{-(1/2)(\boldsymbol{Y}_i - \boldsymbol{f}_i)^T\boldsymbol{R}_i(\boldsymbol{Y}_i - \boldsymbol{f}_i) - (1/2)\widehat{\boldsymbol{b}}_i^T\boldsymbol{D}^{-1}\widehat{\boldsymbol{b}}_i\}.\end{aligned}$$

Thus, to show the equivalence, we first want to show that

$$|\boldsymbol{R}_i||\boldsymbol{D}||\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i| = |\boldsymbol{R}_i + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i|. \tag{1}$$

This is straightforward by using standard matrix results. One way is to invoke this one: If $\boldsymbol{A}$ is $(p \times q)$ and $\boldsymbol{B}$ is $(q \times p)$, then

$$|\boldsymbol{I}_p + \boldsymbol{AB}| = |\boldsymbol{I}_q + \boldsymbol{BA}|.$$

We apply this result to

$$|\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i| = |\boldsymbol{D}^{-1}(\boldsymbol{I} + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)| = |\boldsymbol{D}|^{-1} |\boldsymbol{I} + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i|.$$

The second term in the last expression is thus equal to

$$|\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1}| = |\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T||\boldsymbol{R}_i|^{-1}.$$

Putting all together, we have

$$|\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i| = |\boldsymbol{D}|^{-1} |\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T||\boldsymbol{R}_i|^{-1}.$$

Thus, from above, we can conclude (1).

Now we need to deal with the term in the exponential. We want to show that

$$(\boldsymbol{Y}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{f}_i) + \widehat{\boldsymbol{b}}_i^T \boldsymbol{D}^{-1} \widehat{\boldsymbol{b}}_i = (\boldsymbol{u}_i - \boldsymbol{f}_i)(\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T)^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i), \tag{2}$$

where we have defined

$$\boldsymbol{u}_i = \boldsymbol{Y}_i + \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i$$

so that $(\boldsymbol{u}_i - \boldsymbol{f}_i) = (\boldsymbol{Y}_i - \boldsymbol{h}_i)$ and $\boldsymbol{h}_i$ is defined on page 322. We can write

$$\widehat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i - \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i),$$

which leads to

$$\widehat{\boldsymbol{b}}_i + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i = \boldsymbol{D}(\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i) \widehat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i)$$

so that we finally obtain

$$\widehat{\boldsymbol{b}}_i = (\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i).$$

Note that we can write the left hand side of (2) as

$$(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - \widehat{\boldsymbol{b}}_i^T \boldsymbol{Z}_i \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i + \widehat{\boldsymbol{b}}_i^T (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1}) \widehat{\boldsymbol{b}}_i.$$

Now simplifying this and inserting the expression above for $\widehat{\boldsymbol{b}}_i$, we obtain

$$\begin{aligned}
&(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) \\
&\quad - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) \\
&+ (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1}) (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i).
\end{aligned}$$

This simplifies further to

$$(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i)$$

which can be rewritten as

$$(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \{\boldsymbol{R}_i^{-1} - \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1}\} (\boldsymbol{u}_i - \boldsymbol{f}_i).$$

Using standard matrix inversion results, the middle term is equal to

$$(\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T)^{-1}.$$

Substituting this gives the desired result.

3. (a) This is a subject-specific generalized linear mixed effects model. Thus, (3) is the individual i-specific probability of having severe arthritis symptoms at time t_{ij} for subject i, so that $\exp(\beta_0 + \beta_1 t_{ij} + b_{1i})$ is the odds and $\beta_0 + \beta_1 t_{ij} + b_{1i}$ is the log odds of severe arthritis for individual i at time t_{ij}. Accordingly, $\beta_0 + b_{1i}$ is the log odds at baseline ($t_{i1} = 0$) so that β_0 is the mean or "typical" value of the log odds in the population of individuals; i.e., under the usual convention, the log odds for the "typical individual" who has the "typical" value of log odds. There is no random effect associated with the change in log odds over time, so that, for all i, the amount by which individual log odds changes from baseline after one time unit is β_1, and β_1 is the "typical" value of this change in log odds. Thus, β_0 and β_1 characterize the probability of severe symptoms at any time t_{ij} for an individual for whom the log odds of severe symptoms at baseline and the change in log odds per unit time are equal to the "typical" or average values in the population.

   (b) This is a population-averaged model. Thus, (5) is a model for the overall population probability of having severe arthritis at time t_{ij} for a randomly chosen individual in the population. Thus, β_0 and β_1 characterize this overall population probability. β_0 is thus the log odds, and $\exp(\beta_0)$ is the odds that a randomly chosen individual from the population will have severe arthritis symptoms at baseline, and β_1 is the change in log odds from baseline that a randomly chosen individual from the population will have severe symptoms at 1 time unit post-baseline and $\exp(\beta_1)$ is the multiplicative factor by which the odds of that a randomly chosen individual from the population will have severe symptoms changes in one time unit.

   (c) On the course web page are two SAS programs with their corresponding log files and output. The first fits (3)-(4), with a single "intercept" random effect, and the second fits (3)-(7) with the population model (8), with "intercept" and "slope" random effects. Several different methods are used as implemented in `proc glimmix`, `proc nlmixed`, and the `glimmix` macro – these include linearization about zero ("MQL" type method), linearization about current estimates of random effects ("PQL" type methods), full Laplace approximation, and full adaptive quadrature with 10 and 20 quadrature points

   For model (3)-(4), here are the results. The first five columns are using `proc glimmix`; the Laplace and quadrature (with 10, 20, and 50 quadrature points) results in columns 3-5 are almost identical to those from `proc nlmixed`. The sixth column shows the results from the `glimmix` macro.

	"MQL"	"PQL"	Laplace	Quad-10	Quad-20	Quad-50	"PQL"
β_0	-0.8112	-0.9259	-1.3233	-1.3043	-1.3055	-1.3055	-1.0836
β_1	-0.0508	-0.0651	-0.0861	-0.0897	-0.0898	-0.0898	-0.0945
β_2	-0.1258	-0.1352	-0.1611	-0.1610	-0.1610	-0.1610	-0.1328
D	1.3827	1.6405	3.3774	3.7441	3.7534	3.7534	3.1159

   The results are not very consistent across methods. The approximate linearization methods MQL and PQL using `proc glimmix` give results that are markedly different from those using the full Laplace approximation or adaptive quadrature. The results for quadrature using 20 and 50 quadrature points are similar, suggesting that a sufficient number of points have been used to achieve good accuracy. 10 quadrature points does not seem to be sufficient. The PQL approximation implemented in the `glimmix` macro seems to do a bit better but is still "off" relative to the results obtained with quadrature.

   Clearly, the linearization approximations are not very accurate. This is a binary response, and it is well-known that "PQL" approximation can lead to nontrivial bias in estimators for

the components of β and D. This is likely what is going on here. I would definitely use adaptive quadrature in this kind of situation; it is trivial to perform and appears to yield good accuracy with sufficient number of quadrature points. The full Laplace approximation, which is equivalent to adaptive quadrature with 1 point, does "okay," but underestimates D and overestimated β_0.

Now consider the fit of the model (3)-(7) with two random effects. Here are the results using `proc glimmix` to implement the full Laplace approximation and quadrature; no entries mean that the algorithm did not converge.

	"MQL"	"PQL"	Laplace	Quad-10	Quad-20	Quad-50	"PQL"
β_0	–	–	-0.9613	-1.1016	-1.1017	-1.1018	-1.1795
β_1	–	–	-0.2783	-0.2174	-0.2174	-0.2174	-0.1406
β_2	–	–	-0.2124	-0.2064	-0.2063	-0.2063	-0.1946
D_{11}	–	–	1.1435	2.4025	2.4023	2.4022	5.4794
D_{12}	–	–	0.3816	0.2430	0.2432	0.2432	-0.5662
D_{22}	–	–	0.1308	0.1082	0.1081	0.1081	0.2884

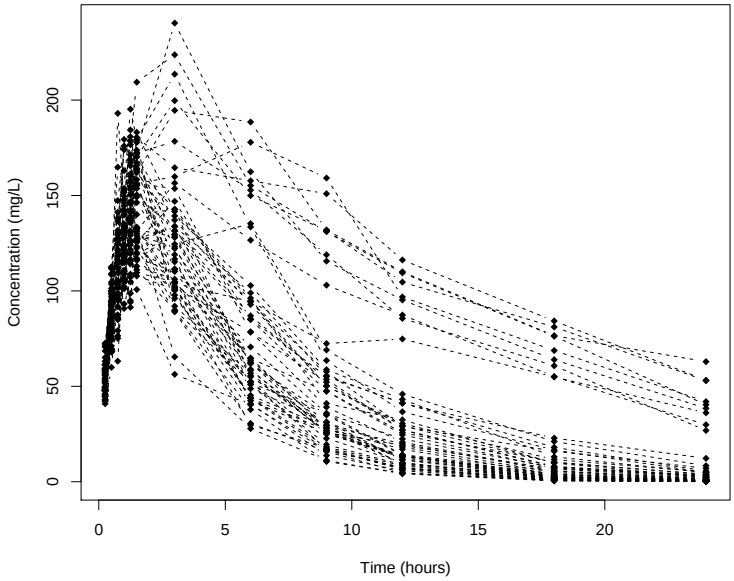
The results using `proc nlmixed` for quadrature are similar but not identical, probably owing to slightly different implementations and perhaps different starting values.

The approximate methods implemented in `proc glimmix` did not converge, and I could not get them to converge by relaxing the convergence criterion. The full Laplace approximation did not converge with the default convergence criteria, but when I relaxed the overall criterion (relative change) from 10^{-8} to 10^{-6} using the `nloptions` statement, it did, although the precision of the results may be suspect. The results from this fit are included above.

The fits using quadrature are all fairly similar, and seem to mostly stabilize with 20 quadrature points, as the results for 20 and 50 are almost identical. However, the fits using the full Lapace approximation and PQL implemented in the `glimmix` macro are different, particular when it comes to the elements of $\boldsymbol{D}$; the latter fit is especially disparate.

(b) The results for estimation of β using `proc glimmix` are $\widehat{\beta} = (-0.8150, -0.05562, -0.1127)^T$. Not surprisingly, these estimates are most similar to those from "MQL" for model (3)-(4), which linearizes about 0 so is essentially fitting the same mean model. They are thus fairly different from those from the most accurate quadrature fit of the subject-specific model (3)-(4). Clearly, the parameters in the population-averaged and subject-specific models represent different quantities with different interpretations.

4. Here is a plot of the data



The individual-specific profiles appear to be well-represented by a one compartment model with first order absorption and elimination. Thus, there are three parameters, fractional rate of absorption k_a, representing absorption characteristics; clearance Cl, representing elimination characteristics; and volume of distribution V, representing distribution characteristics. Because the distributions of these parameters are known to be skewed, I parameterized the model in terms of $\beta_1 = \log k_a$, $\beta_2 = \log Cl$, and $\beta_3 = \log V$. I adopted a subject-specific nonlinear mixed effects model in which these parameters vary across individuals (so each has an associated random effect). I also assumed the power model for within-individual variance and allowed the power δ to be estimated. Because there are enough time points on each individual to fit the one compartment model separately to each, it would be possible to do this, obtain residuals and predicted values for each individual, pool these across individuals, and plot residuals against predicted values to get a visual impression of the nature of within-individual variance. You may have done this.

The investigators are interested in characterizing the mean values of k_a, Cl, and V and whether or not these means are systematic functions of subject characteristics (age, weight, and CYP2D6 phenotype here). The model is parameterized with these on the log scale; for the purpose of investigating associations, it is reasonable to operate on the transformed scale and consider the evidence supporting whether or not the mean values of these on the transformed scales are associated with these covariates. This is the standard approach to this sort of question.

I first fit this general model with a second stage model involving no individual-specific covariates. Based on the hint, I fit all models using at least one of the SAS `nlinmix` macro and SAS `proc nlmixed` (using the full Laplace approximation; I had trouble getting quadrature fits to converge), and R `nlme()`. See the attached SAS and R programs, which show very similar results for the fits of this basic model.

From this basic fit, I obtained the empirical Bayes estimates of the random effects and plotted them against the covariates (box plots or scatter plots as appropriate, not shown). These

plots suggest possible associations of each PK parameter and in particular a strong relationship between CYP2D6 and clearance. I took the evidence in these plots together with a fit of a "full blown" model with all covariates related to all 3 PK parameters to arrive informally at a final model.

With all the stuff being estimated, some of the estimates are probably not very precise, resulting in "high" p-values, but some strong associations are evident. You may have tried fitting several models before settling on a final model. The important associations the investigators would like to know about then follow from what you included in your final model.

Although the estimates of the fixed effects I obtained using the `nlinmix` macro, `proc nlmixed`, and `nlme()` were similar, the estimates of the among-individual covariance matrix $\boldsymbol{D}$ were very different. In particular, those using the SAS implementations were very reasonable, with moderate correlations among elements. However, those obtained using `nlme()` were extremely sensitive to starting values for the fixed effects, and moreover seemed to be driven to having the correlation between log absorption rate and log clearance being almost 1 if not 1. I found this to be very suspicious and disconcerting; I suspect it is not a real phenomenon but some sort of computational artifact. This is why I say it is important to use several rounds of starting values and to compare across implementations. If I were reporting results to investigators, I would feel much more comfortable using those coming out of the SAS implementations.

Based on the final model, the investigators are interested in the average values of k_a, Cl and V in the population and if these are systematically associated with subject characteristics. There are lots of ways to interpret and address these questions. The models here are parameterized with these parameters on the log scale. One possibility is to just report average or "typical values" on the log scale, along with standard errors. So, for example, for a model with $\beta_{2i} = \log Cl_i$ depending on weight and age, e.g.,

$$\beta_{2i} = \beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age} + b_{2i},$$

one might calculate the sample averages of log weight and age and report the estimated average value of β_{2i},

$$\beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age},$$

evaluated at these. Or provide estimates for a range of values of weight and age. Of course, the investigators would probably like estimates of these quantities on their original scales. In the above model, we thus want the expectation of $Cl_i = \exp(\beta_{2i})$,

$$\begin{aligned} E(Cl_i) &= E\{\exp(\beta_{2i})\} = \exp\{\beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age}\}E\{\exp(b_{2i})\} \\ &= \exp\{\beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age}\}\exp(D_{22}/2), \end{aligned}$$

using the fact that $b_{2i} \sim \mathcal{N}(0, D_{22})$. An estimator is thus found by substituting the estimators for β_2 and D_{22}.

If we report an estimate, we'd better report a *standard error* to provide an assessment of the quality of the estimation procedure. To obtain a standard error to go along with this estimate, given the approximate large sample joint covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{D}}$, one can use the *delta method* applied to this and similar expressions to obtain an approximate standard error. (Under normality of everything, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{D}}$ are independent.) Unfortunately, as we have discussed in class, the large sample approximate sampling distribution of $\widehat{\boldsymbol{D}}$ is not very reliable, and these are cumbersome to obtain from the software (if it provides them at all

– `nlme()` has the `apVar` attribute that produces a large sample covariance matrix for the estimators of all the covariance and variance parameters in the model, but it is difficult to interpret. It is possible to get this out of `nlinmix` and `nlmixed` with some wrangling for the former).

It is thus customary to do one of two things. One is to pretend that the covariance parameters are "known" and treat the estimates as fixed, and apply the delta method to expressions like that above only in the elements of $\widehat{\boldsymbol{\beta}}$. The other and most widely adopted is as follows. By a linear Taylor series about $b_{2i} = 0$ (its mean),

$$Cl_i \approx \exp\{\beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age}\} + \exp\{\beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age}\}b_{2i},$$

which implies $E(Cl_i) \approx \exp\{\beta_{20} + \beta_{21}\text{weight} + \beta_{22}\text{age}\}$. Whether it is made explicit or not, this approximation to the "typical" clearance rate is what would ordinarily be done in practice and is the standard approximation used in PK and in interpretations in generalized linear mixed effects models as in the discussion of the epileptic seizure example in Chapter 9. One can then use the delta method to get standard errors based on the large sample covariance matrix of $\widehat{\boldsymbol{\beta}}$.

In the program, we demonstrate implementation of this approximation. We can extract the approximate covariance matrix for the sampling distribution of $\widehat{\boldsymbol{\beta}}$ as shown in the program. For parameters on the log scale, let $\boldsymbol{a}$ be a vector of the same length as $\widehat{\boldsymbol{\beta}}$ such that the quantity of interest can be represented as $\exp(\boldsymbol{a}^T\widehat{\boldsymbol{\beta}})$. Then the delta method proceeds by a linear Taylor series in this quantity in $\widehat{\boldsymbol{\beta}}$ about the "true value" of $\boldsymbol{\beta}$ and then substituting the estimate $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$, we have that $\text{var}\{\exp(\boldsymbol{a}^T\widehat{\boldsymbol{\beta}})\}$ can be approximated/estimated by

$$\exp(2\boldsymbol{a}^T\beta)\,\{\boldsymbol{a}^T\widehat{\text{var}}(\widehat{\boldsymbol{\beta}})\boldsymbol{a}\},$$

where $\widehat{\text{var}}(\widehat{\boldsymbol{\beta}})$ is the estimated large sample covariance matrix of $\widehat{\boldsymbol{\beta}}$. An example calculation is shown in the R program.

You may have done something different. The important thing is that you recognized that you needed to do something to report at least approxmate estimates and standard errors.