ST 790, Homework 4 Solutions
Spring 2018

1. (a) As in the problem statement, we suppress conditioning on $\tilde{\boldsymbol{x}}$. Writing

$$\beta = \beta^* + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}),$$

where $\boldsymbol{u}$ is independent of $\boldsymbol{b}$ and $\boldsymbol{e}$, we can write

$$\boldsymbol{Y} = \boldsymbol{X}\beta^* + \boldsymbol{X}\boldsymbol{u} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{e},$$

so that $E(\boldsymbol{Y}) = \boldsymbol{X}\beta^*$, and it is immediate using the independence of $\boldsymbol{u}$, $\boldsymbol{b}$, and$\boldsymbol{e}$ that

$$\text{var}(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \boldsymbol{Z}\widetilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{R}.$$

It is then straightforward to find the covariances between $\boldsymbol{Y}$ and β, $\boldsymbol{Y}$ and $\boldsymbol{b}$, and $\boldsymbol{b}$ and β; namely

$$\text{cov}(\boldsymbol{Y}, \beta) = \boldsymbol{X}\boldsymbol{H},$$
$$\text{cov}(\boldsymbol{Y}, \boldsymbol{b}) = \boldsymbol{Z}\widetilde{\boldsymbol{D}},$$
$$\text{cov}(\boldsymbol{b}, \beta) = \boldsymbol{0}.$$

It then follows that the joint normal distribution is

$$\mathcal{N}\left\{ \begin{pmatrix} \beta^* \\ \boldsymbol{0} \\ \boldsymbol{X}\beta^* \end{pmatrix}, \begin{pmatrix} \boldsymbol{H} & \boldsymbol{0} & \boldsymbol{H}\boldsymbol{X}^T \\ \boldsymbol{0} & \widetilde{\boldsymbol{D}} & \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T \\ \boldsymbol{X}\boldsymbol{H} & \boldsymbol{Z}\widetilde{\boldsymbol{D}} & \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \boldsymbol{Z}\widetilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{R} \end{pmatrix} \right\}. \tag{1}$$

(b) We can take as the starting point that the joint distribution of distribution of β, $\boldsymbol{b}$, and $\boldsymbol{Y}$ is as in (1). It follows from this that the joint distribution of β and $\boldsymbol{Y}$ is

$$\mathcal{N}\left\{ \begin{pmatrix} \beta^* \\ \boldsymbol{X}\beta^* \end{pmatrix}, \begin{pmatrix} \boldsymbol{H} & \boldsymbol{H}\boldsymbol{X}^T \\ \boldsymbol{X}\boldsymbol{H} & \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \boldsymbol{Z}\widetilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{R} \end{pmatrix} \right\}. \tag{2}$$

We can thus use standard results for the conditional distribution of a partitioned multivariate normal random vector to write down $E(\beta|\boldsymbol{Y})$ and $\text{var}(\beta|\boldsymbol{Y})$. The distribution of $\beta|\boldsymbol{Y}$ is then normal with these moments, and we can then show that these moments can be expressed in the form given in the problem statement.

Take $E(\beta|\boldsymbol{Y})$ first. With $\boldsymbol{V} = \boldsymbol{Z}\widetilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{R}$, we have

$$\begin{aligned} E(\beta|\boldsymbol{Y} = \boldsymbol{y}) &= \beta^* + \boldsymbol{H}\boldsymbol{X}^T(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta^*) \\ &= \{\boldsymbol{I} - \boldsymbol{H}\boldsymbol{X}^T(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\beta^* + \boldsymbol{H}\boldsymbol{X}^T(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}\boldsymbol{y} \\ &= \boldsymbol{A}_1 + \boldsymbol{A}_2. \end{aligned}$$

Consider each term on the right hand side. We evaluate each term using brute force linear algebra. Using standard matrix inversion results in Appendix A repeatedly, we can write the

first term as

$$\begin{aligned}
\boldsymbol{A}_1 &= [\boldsymbol{I} - \boldsymbol{H}\boldsymbol{X}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{X}]\beta^* \\
&= [\boldsymbol{I} - \boldsymbol{H}\{\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}]\beta^* \\
&= [\boldsymbol{I} - \boldsymbol{H}\{(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} + \boldsymbol{H}\}^{-1}]\beta^* \\
&= \{(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} + \boldsymbol{H} - \boldsymbol{H}\}\{(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} + \boldsymbol{H}\}^{-1}\beta^* \\
&= (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\{\boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{H}^{-1}\}\beta^* \\
&= (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\{\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{H}^{-1}\beta^* \\
&= (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{H}^{-1} = \boldsymbol{C}^{-1}\boldsymbol{H}^{-1}\beta^*,
\end{aligned}$$

where $\boldsymbol{C} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1}$ as in the problem statement. We also have

$$\begin{aligned}
\boldsymbol{A}_2 &= \boldsymbol{H}\boldsymbol{X}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{y} \\
&= \{\boldsymbol{H} - \boldsymbol{H}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \\
&= \{\boldsymbol{H}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}) - \boldsymbol{H}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \\
&= (\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} = \boldsymbol{C}^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}.
\end{aligned}$$

Combining these two expressions yields the result that

$$E(\beta|\boldsymbol{Y} = \boldsymbol{y}) = \boldsymbol{C}^{-1}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} + \boldsymbol{H}^{-1}\beta^*).$$

Likewise,

$$\begin{aligned}
\text{var}(\beta|\boldsymbol{Y} = \boldsymbol{y}) &= \boldsymbol{H} - \boldsymbol{H}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \boldsymbol{V})^{-1}\boldsymbol{X}\boldsymbol{H} \\
&= [\boldsymbol{I} - \boldsymbol{H}\boldsymbol{X}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{X}]\boldsymbol{H}.
\end{aligned}$$

By the same calculations as for $\boldsymbol{A}_1$ above, it follows that

$$\text{var}(\beta|bY = \boldsymbol{y}) = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{H}^{-1}\boldsymbol{H} = \boldsymbol{C}^{-1},$$

demonstrating the result.

Note then that the Bayesian estimator for β under this particular choice of prior is

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} + (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{H}^{-1}\beta^*.$$

This evidently has the form of a weighted average of the data and the prior mean for β.

(c) Taking again as the starting point that the joint distribution of distribution of β, $\boldsymbol{b}$, and $\boldsymbol{Y}$ is as in (1), the joint distribution of $\boldsymbol{b}$ and $\boldsymbol{Y}$ is

$$\mathcal{N}\left\{\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{D} & \boldsymbol{D}\boldsymbol{Z}^T \\ \boldsymbol{Z}\boldsymbol{D} & \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \boldsymbol{Z}\widetilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{R} \end{pmatrix}\right\}. \tag{3}$$

As in (b), using standard results for the conditional distribution of a partitioned multivariate normal random vector, we can find $E(\boldsymbol{b}|\boldsymbol{Y})$ and $\text{var}(\boldsymbol{b}|\boldsymbol{Y})$, and the posterior distribution of $\boldsymbol{b}$ is then normal with these moments. From these results, we have

$$\begin{aligned}
&E(\boldsymbol{b}|\boldsymbol{Y} = \boldsymbol{y}) = \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta^*) \\
&\quad = \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta^*)
\end{aligned}$$

$$
\begin{aligned}
&= \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}(\boldsymbol{y} - \boldsymbol{X}\beta^*) \\
&= \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\{\boldsymbol{y} - \boldsymbol{X}\beta^* - \boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} + \boldsymbol{X}\beta^* - \boldsymbol{X}(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\beta^*\} \\
&= \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\Big[\boldsymbol{y} - \boldsymbol{X}\{(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})\beta^* + (\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\beta^* \\
&\qquad - (\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}\Big] \\
&= \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\{\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}^{-1}\beta^* - (\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}\} \\
&= \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\Big[\boldsymbol{y} - \boldsymbol{X}\{\boldsymbol{C}^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} + \boldsymbol{C}^{-1}\boldsymbol{H}^{-1}\beta^*\}\Big].
\end{aligned}
$$

Thus, we have expressed $E(b|\boldsymbol{Y} = \boldsymbol{y})$ in terms of $E(\beta|\boldsymbol{Y})$ in (b).

Similarly, from the results on the conditional distributions of components of a multivariate normal,

$$
\begin{aligned}
\text{var}(\boldsymbol{b}|\boldsymbol{Y} = \boldsymbol{y}) &= \widetilde{\boldsymbol{D}} - \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T)^{-1}\boldsymbol{Z}\widetilde{\boldsymbol{D}} \\
&= \widetilde{\boldsymbol{D}} - \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{H}^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{Z}\widetilde{\boldsymbol{D}} \\
&= \widetilde{\boldsymbol{D}} - \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{C}^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{Z}\widetilde{\boldsymbol{D}}.
\end{aligned}
$$

(d) If we take $\boldsymbol{H}^{-1} = \boldsymbol{0}$ in (b), the posterior distribution of β given $\boldsymbol{Y}$ evaluated at $\boldsymbol{Y} = \boldsymbol{y}$ reduces to a normal with mean

$$
\widehat{\beta} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}
$$

and covariance matrix

$$
(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}.
$$

Thus, under a noninformative prior, the Bayesian estimator for β (the posterior mean/mode) is exactly the maximum likelihood estimator under normality, and the covariance matrix of the posterior is exactly equal to the large sample covariance matrix of that estimator. This is an example of a more general phenomenon that holds in "nice" problems, wherein under a non-informative prior the MLE and Bayesian estimator coincide, and there is a correspondence between the posterior and the large sample distribution.

(e) If we take $\boldsymbol{H}^{-1} = \boldsymbol{0}$ in (c), the posterior distribution of $\boldsymbol{b}$ given $\boldsymbol{Y}$ evaluated at $\boldsymbol{Y} = \boldsymbol{y}$ reduces to a normal with mean

$$
\widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\{\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}\} = \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\{\boldsymbol{y} - \boldsymbol{X}\widehat{\beta}) = \widehat{\boldsymbol{b}}.
$$

From the definitions of the stacked notation on page 173 of the notes, it is straightforward to deduce that this implies that the posterior mean of $\boldsymbol{b}_i$ given $\boldsymbol{Y} = \boldsymbol{y}$, which is the same as that for $\boldsymbol{b}_i$ given $\boldsymbol{Y}_i = \boldsymbol{y}_i$ by independence across i, is given by

$$
\widehat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\beta}),
$$

which is (6.56).

Similarly, with $\boldsymbol{H}^{-1} = \boldsymbol{0}$, we obtain

$$
\text{var}(\boldsymbol{b}|\boldsymbol{Y} = \boldsymbol{y}) = \widetilde{\boldsymbol{D}} - \widetilde{\boldsymbol{D}}\boldsymbol{Z}^T\{\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{Z}\widetilde{\boldsymbol{D}}.
$$

Again from the definitions of the stacked notation on page 173 and independence across i, this implies that

$$\text{var}(\boldsymbol{b}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i) = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i^T\{\boldsymbol{V}_i^{-1} - \boldsymbol{V}_i^{-1}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\}\boldsymbol{Z}_i\boldsymbol{D},$$

which is equal to $\text{var}(\widehat{\boldsymbol{b}}_i)$ given in (6.57).

These results show that the usual BLUP $\widehat{\boldsymbol{b}}_i$ when $\boldsymbol{\xi}$ is known is the Bayesian "estimator" for $\boldsymbol{b}_i$ under a noninformative prior. Likewise, the posterior covariance is the same as the variance of $\widehat{\boldsymbol{b}}_i$ calculated from a frequentist perspective. Thus, the BLUP has a Bayesian interpretation in this sense.

2. *Hepatitis C dynamics for a single subject.* (a) See attached programs and plot of the data with the fitted model superimposed.

   (b) The estimators of the HCV dynamic parameters are $\widehat{V}_0 = \exp(\widehat{\beta}_1)$, $\widehat{c} = \exp(\widehat{\beta}_2)$, and $\widehat{\epsilon} = \exp(\widehat{\beta}_3)/\{1+\exp(\widehat{\beta}_3)\}$. Given we have approximate standard errors for the components of $\widehat{\beta}$ based on the asymptotic theory, an obvious approach is to use the delta method. Generically, if we let $a(\beta)$ be a real-valued function of β, then by the usual linear Taylor series

   $$a(\widehat{\beta}) \approx a(\beta_0) + a_\beta^T(\beta_0)(\widehat{\beta} - \beta_0),$$

   where $a_\beta(\beta)$ is the vector of partial derivatives of $a(\beta)$ with respect to β, so that

   $$a(\widehat{\beta}) \dot{\sim} \mathcal{N}\{a(\widehat{\beta}), a_\beta^T(\widehat{\beta})\widehat{\boldsymbol{\Sigma}}a_\beta(\widehat{\beta})\},$$

   where $\widehat{\boldsymbol{\Sigma}}$ is the large sample approximate covariance matrix for $\widehat{\beta}$. Taking $a(\beta) = \exp(\beta_k)$ for $k = 1, 2$, $a_\beta(\beta_0)$ has $\exp(\beta_{k0})$ in the kth position and zeroes elsewhere. Similarly, taking $a(\beta) = \exp(\widehat{\beta}_3)/\{1+\exp(\widehat{\beta}_3)\}$, $a_\beta(\beta_0)$ has $a_\beta(\beta_0)\{1 - a_\beta(\beta_0)\}$ in the third position and zeroes elsewhere. Then the delta method variances can be obtained from

   $$a_\beta^T(\widehat{\beta})\widehat{\boldsymbol{\Sigma}}a_\beta(\widehat{\beta})$$

   and approximate standard errors found by taking the square root. See the R program for the results.

3. (a) See the R program and SAS output; except in the case of the pretest measures for placebo, the sample variance is substantially larger than the sample mean at both baseline and follow-up. This suggests that the usual Poisson mean-variance relationship does not hold; apparently, there is substantial overdispersion.

   (b) Given that there are only $n = 2$ observations on each individual, the only possible correlation model is

   $$\boldsymbol{\Gamma}_i(\alpha) = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$$

   Letting $\boldsymbol{T}_i(\beta, \sigma^2, \boldsymbol{x}_i)$ be the (2×2) diagonal matrix with diagonal elements

   $$\sigma^2\exp(\beta_1), \sigma^2\exp(\beta_1 + \beta_2 + \beta_3\delta_i^{(1)} + \beta_4\delta_i^{(2)}),$$

   the covariance matrix is

   $$\boldsymbol{T}^{1/2}(\beta, \sigma^2, \boldsymbol{x}_i)\boldsymbol{\Gamma}_i(\alpha)\boldsymbol{T}^{1/2}(\beta, \sigma^2, \boldsymbol{x}_i).$$

(c) For brevity, write f_{i1} to denote $f(\boldsymbol{x}_i, \beta)$ evaluated at $t_{i1} = 0$, which is equal to $\exp(\beta_1)$, and f_{i2} to denote $f(\boldsymbol{x}_i, \beta)$ evaluated at $t_{i2} = 1$, which is equal to $\exp(\beta_1 + \beta_2 + \beta_3\delta_i^{(1)} + \beta_4\delta_i^{(2)})$. The corresponding variances are $\sigma^2 g_{ij}^2 = \sigma^2 f_{ij}$ at t_{ij}, $j = 1, 2$. Because we have both the variance scale parameter σ^2 and the scalar correlation parameter α, $\boldsymbol{\xi} = (\sigma^2, \alpha)^T$, to estimate $\boldsymbol{\xi}$, there will be 3 terms in $\boldsymbol{u}_i$, two squared terms and one cross-product term. Placing these in the order on page 242, it follows that

$$\boldsymbol{u}_i = \begin{pmatrix} (Y_{i1} - f_{i1})^2 \\ (Y_{i1} - f_{i1})(Y_{i2} - f_{i2}) \\ (Y_{i2} - f_{i2})^2 \end{pmatrix}, \quad \boldsymbol{v}_i = \begin{pmatrix} \sigma^2 g_{i1}^2 \\ \sigma^2 \alpha g_{i1} g_{i2} \\ \sigma^2 g_{i2}^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 f_{i1} \\ \sigma^2 \alpha f_{i1}^{1/2} f_{i2}^{1/2} \\ \sigma^2 f_{i2} \end{pmatrix},$$

$$\boldsymbol{E}_i = \begin{pmatrix} 2\sigma g_{i1}^2 & 0 \\ 2\sigma\alpha g_{i1} g_{i2} & \sigma^2 g_{i1} g_{i2} \\ 2\sigma g_{i2}^2 & 0 \end{pmatrix} = \begin{pmatrix} 2\sigma f_{i1} & 0 \\ 2\sigma\alpha f_{i1}^{1/2} f_{i2}^{1/2} & \sigma^2 f_{i1}^{1/2} f_{i2}^{1/2} \\ 2\sigma f_{i2} & 0 \end{pmatrix}.$$

To deduce the "covariance matrix" $\boldsymbol{Z}_i(\beta, \boldsymbol{\xi})$ under the "Gaussian working assumption," we need to use the definition of $\boldsymbol{u}_i$ above along with the condition (8.23) that holds under normality. In general, (8.23) for $\boldsymbol{u}_i$ defined as above implies that

$$\boldsymbol{Z}_i = \sigma^4 \begin{pmatrix} 2g_{i1}^4 & 2\alpha g_{i1}^3 g_{i2} & 2\alpha^2 g_{i1}^2 g_{i2}^2 \\ & (1 + \alpha^2) g_{i1}^2 g_{i2}^2 & 2\alpha g_{i1} g_{i2}^3 \\ & & 2g_{i2}^4 \end{pmatrix}.$$

This can be rewritten in terms of f_{i1} and f_{i2}.

(c) I've omitted a plot of the data, but you may have made one. The means for each treatment at each time point in (a) mostly tell the story. The model (7) is a fully saturated model in that it represents the mean for each of the two time points for each treatment separately (there is an assumed common mean at baseline based on the randomization, although the sample means seem to suggest that the baseline means may be different by treatment).

The attached programs show different fits of the mean-variance model. Because there are only two time points, there is only a single correlation, and we can incorporate the correlation model in (b) by specifying an unstructured correlation, compound symmetry, or AR(1), and all should lead to the same result. `proc genmod`, `proc gee`, and the `gee()` function all give the same model fit when the working correlation is taken to be compound symmetric. However, when it is taken to be unstructured or AR(1) in `gee()` or when `geeglm()` is used, the fit is different and is the same as that obtained from `proc glimmix` using the quadratic estimating equation.

Regardless of the fit, there seems to be strong evidence that both antibiotics lead to smaller mean leprosy bacilli count at the follow-up time relative to placebo; the estimates of β_3 and β_4 are negative with associated small p-values. You might have done a joint test of $\beta_3 = \beta_4 = 0$; see the SAS program for this test conducted in the call to `proc genmod` with compound symmetry. `proc glimmix` creates this test automatically for this parameterization of the model.

4. A plot of the raw data is not that informative, but we can summarize the proportions of subjects with severe symptoms at each month under each of placebo and active drug.

month	0	2	4	6
placebo	0.352	0.297	0.276	0.255
auranofin	0.324	0.124	0.200	0.145

The raw proportions suggest that probability of severe symptoms status stays relatively flat over the study period in the placebo group, while in the active drug group it shows a variable but downward trend. You may have plotted these proportions to get a visual impression.

These observations suggest that the following model might be a reasonable basic framework in which to address the questions; you might have done something different. Let Y_{ij} be Arthritis Self-Assessment for subject i at month t_{ij}, where $t_{ij} = 0, 2, 4, 6$ months for $j = 1, \ldots, 4$ for all subjects and $i = 1, \ldots, 290$. Let $\delta_i = 0$ if i was assigned to placebo and 1 if assigned to auranofin therapy. Here, there are no within-individual covariates, so that $\boldsymbol{z}_{ij} = t_{ij}$, and the among-individual covariates $\boldsymbol{a}_i = (\delta_i, a_i)$, where a_i is the age of subject i at baseline, so that $\boldsymbol{x}_{ij} = (t_{ij}, \delta_i)^T$ and $\boldsymbol{x}_i = (t_{i1}, \ldots, t_{in_i}, \delta_i, a_i)^T$.

We adopt the following basic model:

$$E(Y_{ij}|\boldsymbol{x}_i) = E(Y_{ij}|\boldsymbol{a}_i) = \frac{\exp(\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}\delta_i)}{1 + \exp(\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}\delta_i)}. \tag{4}$$

In (4), the probability of having severe symptoms at baseline is $e^{\beta_0}/(1 + e^{\beta_0})$ in both groups, which makes sense given that this is a randomized study. In the placebo group, this probability is $e^{\beta_0+\beta_1 t}/(1 + e^{\beta_0+\beta_1 t})$ at month t post-baseline, and for the active drug it is $e^{\beta_0+\beta_1 t+\beta_2 t}/(1 + e^{\beta_0+\beta_1 t+\beta_2 t})$. Thus, $\beta_1 t$ is the change in log odds of severe symptoms from baseline at time t for placebo patients, and $(\beta_1 + \beta_2)t$ is the change in log odds from baseline at time t for active drug patients. If β_1 is different from zero, then the probability of severe symptoms changes over the study period under placebo. If $\beta_1 + \beta_2$ is different from zero, then the probability of severe symptoms changes over the study period under auranofin, and $\beta_2 t$ represents the difference in log odds at time t post-baseline between the two groups. Clearly, then, if $\beta_2 = 0$, the active drug has no effect, and the probability of severe symptoms under placebo and auranofin is the same at any time t.

The attached programs implement fits of model (4); these lead to slightly different estimates in some cases between SAS and R implementations and using moment-based (`proc genmod`, `gee()`, `geeglm()`) and quadratic estimating equations (`proc glimmix`). Because these data are miraculously balanced, it is possible to fit an unstructured working correlation model. The estimated correlation matrix suggests that a working compound symmetric correlation structure is reasonable. Thus, a sensible approach is to adopt the working model and use the robust sandwich standard errors "just in case." Alternatively, as the informal evidence is fairly supportive of an approximate compound symmetric structure, you may have chosen to use model-based standard errors. Either is fine, as long as it is explicitly justified. Also fitted is (4) in the alternative parameterization

$$E(Y_{ij}|\boldsymbol{x}_i) = E(Y_{ij}|\boldsymbol{a}_i) = \frac{\exp(\beta_0 + \beta_1 t_{ij}(1 - \delta_i) + \beta_2 t_{ij}\delta_i)}{1 + \exp(\beta_0 + \beta_1 t_{ij}(1 - \delta_i) + \beta_2 t_{ij}\delta_i)}, \tag{5}$$

which allows us to address (i) directly. From the fits of (5), it is immediate that there is strong evidence that the probability of severe symptoms changes over the study period for auranofin, and the estimate of β_2 is negative, suggesting that this probability decreases. The evidence is inconclusive for placebo; the estimate of β_1 is negative, which is consistent with the sample proportions (which seem to decrease slightly over the study period), but the standard tests fail to reject the null hypothesis that $\beta_1 = 0$, although the p-values are not too large. Thus, the evidence seems to suggest that auranofin leads to a lower probability of severe symptoms than placebo over the 6 month study period. From the fit of (4), $\beta_2 = 0$

indicates no difference in the probability between placebo and auranofin; from fits of this model, there appears to be strong evidence that the probability is different, consistent with this.

Question (ii) also asks for estimates of the probabilities of severe symptoms at 6 months for each treatment. These can be obtained by substitution in the fitted model, and standard errors can be obtained using the delta method. See the R program.

To address (iii) and (iv), you may have added additional terms allowing the log odds to depend on age at baseline, both at baseline and in the way these change over the study period. See the attached programs for representative fits of models including age. The evidence is not strong enough to suggest associations between age either at baseline and over the course of the study; see the SAS program output.

The conclusion seems to be that the active drug, auranofin, leads to lower probability of severe symptoms than placebo, but there is not enough evidence to suggest that this effect is associated with age.