The following scenario is based on a real study, though most details have been changed and the data provided are not real. Attached to this question is a set of R output that you will want to consult at appropriate points in answering the following questions about this scenario and the analysis of data from it.

An experiment is done to determine whether the progressive effects of a particular disease can be detected in an electrical response of rat eyes to flashes of light. Four animals (rats) are infected with the disease and

$$y = \text{a particular electrical response to a "standard" light flash } (\mu\text{V})$$

Two measurements are made on each infected rat at one month, two months, and then three months after infection, for a total of 6 observations per infected rat. To account for the fact that (unintended) changes in lab conditions and/or measurement equipment might occur over time, 2 measurements were also made on 2 uninfected animals at each measurement period. But, these were different animals each month. Thus there were a total of $4+2+2+2=10$ rats involved in the study. Four of these were infected and observed at multiple periods and 6 were control animals that were observed at only a single point in time.

The questions of greatest scientific interest in this study are questions such as "Is there a detectable difference in electrical response between rats that have been infected and those that have not?" and "Can changes in electrical response over time be detected for infected rats?"

1) In light of the basic goals of the study, discuss in qualitative terms any additional utility that would have been provided had the study

- **a)** included measurement of the infected animals at month 0 (just before infection).
- **b)** included measurement of each control animal at all of months $0,1,2,$ and 3.

Label the infected rats $i=1,2,3,4$ and the control rats $i=5,6,\ldots,9,10$. Suppose control rats 5 and 6 are tested at month 1, rats 7 and 8 are tested at month 2, and rats 9 and 10 are tested at month 3. Let

$$y_{ijk} = \text{the } k\text{th response measured on rat } i \text{ at period } j$$

For

$$\mu_0 = \text{a mean response for an uninfected rat}$$
$$\mu_1 = \text{a month 1 mean repsonse for an infected rat}$$
$$\mu_2 = \text{a month 2 mean repsonse for an infected rat}$$
$$\mu_3 = \text{a month 3 mean repsonse for an infected rat}$$

and

$$r_i = \text{a rat effect for rat } i, \quad i=1,2,\ldots,10$$
$$m_j = \text{a month effect for month } j, \quad j=1,2,3$$

we will initially consider models for this situation of the basic form

$$y_{ijk} = \mu_0 I[i>4] + \mu_j I[i \le 4] + r_i + m_j + \varepsilon_{ijk} \qquad (1)$$

for all $n=36$ relevant combinations of $i, j,$ and k.

First, for the sake of simplicity, consider only the *first* of the two replicate measurements made on only rats $1,2,5,7,$ and 9 at each time period, and an ordinary (fixed effects) Gauss-Markov linear model version of (1). Let

$$\mathbf{Y}' = \left(y_{111}, y_{121}, y_{131}, y_{211}, y_{221}, y_{231}, y_{511}, y_{721}, y_{931}\right)$$

and

$$\boldsymbol{\beta}' = \left(\mu_0, \mu_1, \mu_2, \mu_3, r_1, r_2, r_5, r_7, r_9, m_1, m_2, m_3\right)$$

2) Find a matrix $\mathbf{X}$ so that model (1) for these observations can be written in the usual linear model form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Is the resulting model of full rank? (Argue this carefully one way or the other.)

3) Is the quantity $\mu_1 - \mu_0$ estimable in this model? (Argue this carefully one way or another.)

It is probably both more reasonable and also more effective in terms of statistical inference to interpret the rat and month effects in equation (1) as random rather than fixed. So **henceforth** suppose that the r_i's in (1) are iid $\mathrm{N}\left(0, \sigma_r^2\right)$ independent of m_j's that are iid $\mathrm{N}\left(0, \sigma_m^2\right)$, all of which are independent of ε_{ijk}'s that are iid $\mathrm{N}\left(0, \sigma^2\right)$.

Continue for the present to consider only the *first* of the two replicate measurements made on only rats $1,2,5,7,$ and 9 at each time period and $\mathbf{Y}$ exactly as listed at the top of this page.

4) Find matrices $\mathbf{X}$ and $\mathbf{Z}$ and vectors $\boldsymbol{\beta}$ and $\mathbf{u}$ so that model (1) for these observations can be written in the usual mixed linear model form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

Is the resulting model of full rank? (Argue this carefully one way or the other.) Are the scientifically interesting quantities $\mu_i - \mu_0$ for $i = 1,2,3$ and $\mu_2 - \mu_1, \mu_3 - \mu_2,$ and $\mu_3 - \mu_1$ all estimable from these data? Why or why not?

Now begin to consider the whole data set from this study.

5) Based on the mixed effects model described by (1) and the distributional assumptions indicated just after **3)** above, evaluate the following in terms of model parameters:

- **a)** $\text{Var}\left(y_{ijk}\right)$ for any set of indices i, j, k in the data set
- **b)** $\text{Corr}\left(y_{111}, y_{112}\right)$
- **c)** $\text{Corr}\left(y_{111}, y_{211}\right)$ and $\text{Corr}\left(y_{111}, y_{121}\right)$
- **d)** $\text{Corr}\left(y_{111}, y_{221}\right)$

6) Each pair $\left(y_{ij1}, y_{ij2}\right)$ in the data set can be used to compute a sample variance, say s_{ij}^2. How does this sample variance compare to

$$\frac{1}{2-1}\sum_{k=1}^{2}\left(\varepsilon_{ijk}-\bar{\varepsilon}_{ij.}\right)^2$$

(the "sample variance" of the pair $\left(\varepsilon_{ij1}, \varepsilon_{ij2}\right)$)? In light of this, what is the mean of the average sample variance, namely

$$\text{E}\left(\frac{1}{18}\sum_{ij} s_{ij}^2\right) \quad ?$$

Using (1) write $\bar{y}_{1..}, \bar{y}_{2..}, \bar{y}_{3..}$, and $\bar{y}_{4..}$ (the simple averages of the rat 1,2,3, and 4 responses) in terms of averages of appropriate fixed and random effects and random errors. Based on this, what is the expected value of the sample variance of these four sample means? What is the expected value of the sample variance of rat 5 and 6 averages? Of rat 7 and 8 averages? Of rat 9 and 10 averages?

Call the sample variances for groups of rat means referred to above by the respective names $s^2(1,2,3,4), s^2(5,6), s^2(7,8)$, and $s^2(9,10)$. What are sufficient conditions on constants $c_0, c_{1234}, c_{56}, c_{78}, c_{9\ 10}$ under which

$$\text{E}\left(c_0\left(\frac{1}{18}\sum_{ij} s_{ij}^2\right)+c_{1234}s^2(1,2,3,4)+c_{56}s^2(5,6)+c_{78}s^2(7,8)+c_{9\ 10}s^2(9,10)\right)=\sigma_r^2$$

7) The current version of the `lmer()` function in the `lme4` package of Bates produces point estimates of variance components (and their square roots) and Bayes credible intervals (presumably based on Jeffreys priors) for the "error" standard deviation in a mixed linear model and ratios of the other model standard deviations to the error standard deviation. Here the credible intervals are

```
$sigma
         lower    upper
[1,] 0.9412178 1.663813
attr(,"Probability")

$ST
     lower     upper
[1,]     0 0.8048939
[2,]     0 1.7889483
attr(,"Probability")
[1] 0.95
```

Discuss what these results tell you about sources of variability in measuring y on rat eyes.

8) Is there clear evidence of any difference in electrical response to light flash between uninfected and infected rat eyes (at any stage of the disease)? Explain carefully. Is there clear evidence that electrical response changes with time for an infected rat? Explain carefully.

9) No rats were actually tested at month 0 (just before rats 1 through 4 were infected). But based on the data in hand, you might predict what rat 1's responses would have been like at month 0. Give a sensible point prediction for a single test result for rat 1 at month 0. Then give a sensible point prediction *and standard error of that prediction* for a single test for an uninfected rat not included in the data set (say rat 11) at month 0.

Let

x_{ij} = the number of months that rat i has been infected when measurements are taken at month j

An alternative to the mixed model version of (1) that assumes that change in electrical response is linear in time since infection, but that different rats have different rates of change (while still allowing for month effects on measurements) is

$$y_{ijk} = \beta_0 + (\beta_1 + r_i) x_{ij} + m_j + \varepsilon_{ijk} \qquad (2)$$

where we will assume that the r_i's in (2) are iid $\mathrm{N}(0, \sigma_r^2)$ independent of m_j's that are iid $\mathrm{N}(0, \sigma_m^2)$, all of which are independent of ε_{ijk}'s that are iid $\mathrm{N}(0, \sigma^2)$.

10) Evaluate $\mathrm{Var}(y_{111})$, $\mathrm{Var}(y_{121})$, and $\mathrm{Corr}(y_{111}, y_{121})$ in terms of parameters of model (2).

11) Find any unbiased estimator of σ_r^2 in model (2).

R Printout

```
> MoreRats
       y Rat Month Mu
1  50.49   1     1  1
2  51.87   1     1  1
3  47.70   1     2  2
4  45.85   1     2  2
5  44.79   1     3  3
6  44.75   1     3  3
7  49.82   2     1  1
8  49.56   2     1  1
9  46.76   2     2  2
10 46.39   2     2  2
11 40.80   2     3  3
12 40.36   2     3  3
13 48.74   3     1  1
14 50.60   3     1  1
15 48.55   3     2  2
16 48.82   3     2  2
17 45.16   3     3  3
18 43.95   3     3  3
19 51.59   4     1  1
20 50.70   4     1  1
21 47.76   4     2  2
22 47.15   4     2  2
23 42.96   4     3  3
24 44.69   4     3  3
25 48.92   5     1  0
26 49.46   5     1  0
27 52.11   6     1  0
28 51.18   6     1  0
29 47.50   7     2  0
30 49.23   7     2  0
31 46.73   8     2  0
32 49.63   8     2  0
33 48.50   9     3  0
34 47.31   9     3  0
35 49.23  10     3  0
36 48.04  10     3  0

> eyeout<-lmer(y~Mu+(1|Rat)+(1|Month))

> summary(eyeout)
Linear mixed model fit by REML
Formula: y ~ Mu + (1 | Rat) + (1 | Month)
   AIC   BIC logLik deviance REMLdev
 130.6 141.7 -58.29    121.8   116.6



Random effects:
 Groups   Name        Variance Std.Dev.
 Rat      (Intercept) 0.65612  0.81001
 Month    (Intercept) 0.89120  0.94403
 Residual             1.26390  1.12423
Number of obs: 36, groups: Rat, 10; Month, 3
```

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  48.9867     0.7154   68.48
Mu1           0.6040     0.8865    0.68
Mu2          -1.1996     0.8865   -1.35
Mu3          -5.1381     0.8865   -5.80

Correlation of Fixed Effects:
    (Intr) Mu1    Mu2
Mu1 -0.339
Mu2 -0.339  0.323
Mu3 -0.339  0.323  0.323

> vcov(eyeout)
4 x 4 Matrix of class "dpoMatrix"
           [,1]       [,2]       [,3]       [,4]
[1,]  0.5117438 -0.2146778 -0.2146778 -0.2146778
[2,] -0.2146778  0.7859343  0.2540876  0.2540876
[3,] -0.2146778  0.2540876  0.7859343  0.2540876
[4,] -0.2146778  0.2540876  0.2540876  0.7859343

> ranef(eyeout)
$Rat
   (Intercept)
1   0.37817047
2  -1.10549417
3   0.42485040
4   0.30247330
5  -0.31951647
6   0.93101854
7  -0.10549049
8  -0.19972633
9  -0.33906738
10  0.03278213

$Month
  (Intercept)
1   0.8305952
2  -0.4145720
3  -0.4160232


> fitted(eyeout)
 [1] 50.79942 50.79942 47.75067 47.75067 43.81067 43.81067 49.31576 49.31576
 [9] 46.26701 46.26701 42.32701 42.32701 50.84610 50.84610 47.79735 47.79735
[17] 43.85735 43.85735 50.72372 50.72372 47.67497 47.67497 43.73497 43.73497
[25] 49.49775 49.49775 50.74828 50.74828 48.46660 48.46660 48.37237 48.37237
[33] 48.23158 48.23158 48.60343 48.60343

> sim<-mcmcsamp(eyeout, 1000000)

> HPDinterval(sim)
$fixef
                lower      upper
(Intercept) 47.400550 50.5627802
Mu1         -1.036940  2.5609806
Mu2         -2.927035  0.3629514
Mu3         -6.871015 -3.5825394
attr(,"Probability")
[1] 0.95
```

```
$ST
     lower     upper
[1,]     0 0.8048939
[2,]     0 1.7889483
attr(,"Probability")
[1] 0.95

$sigma
         lower    upper
[1,] 0.9412178 1.663813
attr(,"Probability")
[1] 0.95
```