# Stat 511 HW#5 Spring 2009

1. The following is a "calibration" scenario met in an engineering lab. Several resistance temperature devices (RTDs) from a large batch of the devices (that can be used to provide cheap temperature measurements in terms of resistance they register) were placed in a water bath with an extremely high quality thermometer (that we will assume produces the "true" temperature of the bath). At some regular intervals, both the "true temperature" and measured resistances were recorded. We'll let

$$y_{ij} = \text{the measured resistance produced by RTD } i \text{ at time } j$$

and suppose that these are linear functions of the true temperatures plus random error. That is, we'll assume that for

$$t_j = \text{the } j\text{th measured temperature}$$

it is appropriate to model as

$$y_{ij} = \alpha_i + \beta_i t_j + \varepsilon_{ij}$$

where the $\alpha_i$ and $\beta_i$ are an intercept and a slope specific to the particular RTD being studied. Further suppose that the $\alpha_i$ and $\beta_i$ can be described as

$$\alpha_i = \alpha + \gamma_i \text{ and } \beta_i = \beta + \delta_i$$

where $\alpha$ and $\beta$ are unknown constants and the $\gamma_i$ and $\delta_i$ are unobservable random effects. We'll assume that all the random effects have mean 0, that is that

$$\mathrm{E}\,\gamma_i = 0 \;\forall i,\; \mathrm{E}\,\delta_i = 0 \;\forall i, \text{ and } \mathrm{E}\,\varepsilon_{ij} = 0 \;\forall i, j$$

We'll further assume that variances are as

$$\mathrm{Var}\,\gamma_i = \sigma_\gamma^2 \;\forall i,\; \mathrm{Var}\,\delta_i = \sigma_\delta^2 \;\forall i, \text{ and } \mathrm{Var}\,\varepsilon_{ij} = \sigma^2 \;\forall i, j$$

and that all of the random effects are uncorrelated (have 0 covariances). We then have a model with the 5 parameters

$$\alpha, \beta, \sigma_\gamma^2, \sigma_\delta^2, \text{ and } \sigma^2$$

The first two of these are "fixed effects" and the last three are "variance components."

This scenario fits into the "Mixed Linear Model" framework introduced in class. For sake of argument, suppose that there are 3 RTDs and only 3 different measured temperatures, with respective (coded) values $t_1 = 0, t_2 = 1,$ and $t_3 = 4$ .

a) Write out in matrix form the mixed linear model for $\mathbf{Y} = (y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}, y_{31}, y_{32}, y_{33})'$ . (What are $\mathbf{X}, \boldsymbol{\beta}, \mathbf{Z},$ and $\mathbf{u}$ ?)

b) What is $\mathrm{E}\,\mathbf{Y}$ ? Write out and simplify as much as you can the covariance matrix, $\mathrm{Var}\,\mathbf{Y}$ . (Do what you can in terms of using multiplication of partitioned matrices to make the form look simple.)

c) Suppose that in fact $\mathbf{Y} = (99.8, 108.1, 136.0, 100.3, 109.5, 137.7, 98.3, 110.1, 142.2)'$ and that I somehow know that $\sigma_\gamma^2 = 1, \sigma_\delta^2 = 1,$ and $\sigma^2 = .25$ . I can then use generalized least squares to estimate the fixed effects vector $(\alpha, \beta)$ . Do this. (Note that this is estimation of the average intercept and slope for calibration of RTDs of this type.) Does the answer change if I know only that $\sigma_\gamma^2 = \sigma_\delta^2 = 4\sigma^2$ ? Explain. (Indeed, can I even *get* a BLUE for $(\alpha, \beta)$ with only this knowledge?)

d) Suppose that I know that $\sigma_\gamma^2 = 1, \sigma_\delta^2 = 1,$ and $\sigma^2 = 1$. Use again the data vector from part c). What is the BLUE of $(\alpha, \beta)$ under these circumstances?

e) Suppose that it is your job to estimate the variance components in this model. One thing you might consider is maximum likelihood under a normal distribution assumption. This involves maximizing the likelihood as a function of the 5 parameters and it's clear how to get the profile likelihood for the variance components alone. That is, for a fixed set of variance components $(\sigma_\gamma^2, \sigma_\delta^2, \sigma^2)$ one knows $\mathrm{Var}\,\mathbf{Y}$, and may use generalized least squares to estimate $\mathrm{E}\,\mathbf{Y}$ and plug that into the likelihood function in order to get the profile likelihood for $(\sigma_\gamma^2, \sigma_\delta^2, \sigma^2)$. Consider the two vectors of variance components used in parts c) and d) of this problem and the data vector from c). Which set has the larger value of profile likelihood (or profile loglikelihood)?

f) Add the `MASS` and `nmle` packages to your `R` session. Then type

```
> I3<-diag(c(1,1,1))
> J3<-matrix(c(1,1,1,1,1,1,1,1,1),3,3)
> M<-matrix(c(0,0,0,0,1,4,0,4,16),3,3)
> X<-matrix(c(rep(1,9),rep(c(0,1,4),3)),9,2)
> Y<-
matrix(c(99.8,108.1,136.0,100.3,109.5,137.7,98.3,110.1,142.2),9,1)
```

Using these, you can create a negative profile loglikelihood function for $(\sigma_\gamma^2, \sigma_\delta^2, \sigma^2)$ as below. (We're using a *negative* profile loglikelihood here because the `R` optimizer seeks to minimize rather than maximize a function.)

```
> minuslLstar<-function(s2,Y)
+ {
+ temp0<-kronecker(I3,((s2[1]*J3)+(s2[2]*M)+(s2[3]*(I3))))
+ temp1<-ginv(temp0)
+ temp2<-X%*%ginv(t(X)%*%temp1%*%X)%*%t(X)%*%temp1%*%Y
+ temp3<-(Y-temp2)
+ (.5*(log(det(temp0))))+
+ (4.5*log(2*pi))+
+ (.5*(t(temp3)%*%temp1%*%temp3))
+ }
```

Evaluate this function for the two sets of variance components in parts c) and d) as below and verify that your answers are consistent with what you got above.

```
> minuslLstar(c(1,1,1),Y)
> minuslLstar(c(1,1,.25),Y)
```

I did some "hunt and peck"/by hand" optimization of this function, and came up with what we'll use as starting values of $\sigma_\gamma^2 = .07, \sigma_\delta^2 = .45,$ and $\sigma^2 = .37$. Use the `R` function `optim` to find better values (MLEs) by typing

```
> optim(c(.07,.45,.37),minuslLstar,Y=Y,hessian=TRUE)
```

This will set in motion an iterative optimization procedure with the starting value above. This call produces a number of interesting summaries, including a matrix of second partials of the objective function evaluated at the optimum (the Hessian). What are the MLEs?

g) Now consider seeking REML estimates of the variance components. Compute the projection matrices $\mathbf{P}_\mathbf{X}$ and $\mathbf{N} = \mathbf{I} - \mathbf{P}_\mathbf{X}$. Notice that since $(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{X} = \mathbf{0}$, every row of $\mathbf{N}$ is (after transposing) in $C(\mathbf{X}')^\perp$ and so any set of 7 rows of $\mathbf{N}$ that make a $7 \times 9$ matrix, say $\mathbf{B}$, of rank 7 will serve to create the vector of "error contrasts" $\mathbf{r} = \mathbf{BY}$ which one uses to find REML estimates here. Verify that the first 7 rows of $\mathbf{N}$ will work in the present situation by typing

```
> B<-rbind(N[1,],N[2,],N[3,],N[4,],N[5,],N[6,],N[7,])
> qr(B)$rank
```

Note that by construction, $\mathbf{r} \sim \text{MVN}_7\left(\mathbf{0}, \mathbf{B}\left(\mathbf{V}\left(\left(\sigma_\gamma^2, \sigma_\delta^2, \sigma^2\right)\right)\right)\mathbf{B}'\right)$ and "REML" is maximum likelihood for the variance components based on $\mathbf{r}$. You may create a negative loglikelihood function here as

```
> minuslLstar2<-function(s2,Y,B)
+ {
+ temp0<-kronecker(I3,((s2[1]*J3)+(s2[2]*M)+(s2[3]*(I3))))
+ temp1<-B%*%temp0%*%t(B)
+ temp2<-ginv(temp1)
+ temp3<-B%*%Y
+ (.5*(log(det(temp1))))+
+ (3.5*log(2*pi))+
+ (.5*(t(temp3)%*%temp2%*%temp3))
+ }
```

Then you may find REML estimates of the variance components by optimizing this function. Use the starting values from part f) and type

```
> optim(c(.07,.45,.37),minuslLstar2,Y=Y,B=B,hessian=TRUE)
```

How do your estimates here compare to the MLEs you found in part f)? (Statistical folklore says that usually REML estimates are bigger than MLEs.)

h) The (unavailable) BLUE of any (vector of) estimable function(s) $\mathbf{C}\boldsymbol{\beta}$ is the generalized least squares estimator $\widehat{\mathbf{C}\boldsymbol{\beta}} = \mathbf{C}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$. Use first maximum likelihood and then REML estimates of the variance components to estimate $\mathbf{V}$ and then produce realizable estimates of the vector of fixed effects $(\alpha, \beta)'$ as $\widehat{\widehat{\mathbf{C}\boldsymbol{\beta}}} = \mathbf{C}\left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}$ for the $\mathbf{C} = \mathbf{I}$ case.

i) As on panel 811 of Koehler's notes, $\widehat{\mathbf{C}\boldsymbol{\beta}}$ has covariance matrix $\mathbf{C}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{C}'$ and so the covariance matrix of $\widehat{\widehat{\mathbf{C}\boldsymbol{\beta}}}$ from part h) might possibly be estimated as $\mathbf{C}\left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-}\mathbf{C}'$. Use this fact to produce *very* approximate standard errors for your estimates of fixed effects in part h).

j) If $\hat{\mathbf{Y}}^*\left(\boldsymbol{\sigma}^2\right)$ is the (unavailable) generalized least squares estimate of the mean of $\mathbf{Y}$, the BLUP of the vector of random effects is

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{GZ'V}^{-1}\left(\mathbf{Y}-\hat{\mathbf{Y}}^*\left(\boldsymbol{\sigma}^2\right)\right) \\ &= \mathbf{GZ'V}^{-1}\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-}\mathbf{X'V}^{-1}\right)\mathbf{Y}\end{aligned}$$

Letting $\mathbf{B}=\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-}\mathbf{X'V}^{-1}$ and $\mathbf{P}=\mathbf{V}^{-1}\left(\mathbf{I}-\mathbf{B}\right)$, this is

$$\hat{\mathbf{u}}=\mathbf{GZ'V}^{-1}\left(\mathbf{I}-\mathbf{B}\right)\mathbf{Y}=\mathbf{GZ'PY}$$

Estimates of variance components lead not only to an estimate of $\mathbf{V}$, but also to estimates of $\mathbf{G}$ and $\mathbf{P}$. Use first maximum likelihood and then REML estimates of variance components to approximate the BLUP as

$$\hat{\hat{\mathbf{u}}}=\widehat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{P}}\mathbf{Y}$$

k) As the BLUP $\hat{\mathbf{u}}$ is an (admittedly somewhat unpleasant) matrix times $\mathbf{Y}$, it is possible to work out a prediction variance for it,

$$\mathrm{Var}\left(\hat{\mathbf{u}}-\mathbf{u}\right)=\mathbf{G}-\mathbf{GZ'PZG}$$

Once again, estimates of variance components lead to *very* approximate standard errors of the available predictions $\hat{\hat{\mathbf{u}}}$ based on diagonal elements of this matrix. Find these using first maximum likelihood and then REML estimates of variance components.

l) The handout on BLUPs discusses prediction of a quantity

$$l=\mathbf{c}'\boldsymbol{\beta}+\mathbf{s}'\mathbf{u}$$

for an estimable $\mathbf{c}'\boldsymbol{\beta}$. In light of that material, consider again the original context of calibration of the RTDs. Use first MLEs of variance components and then REML estimates to find predictors for

$$\alpha_1=\alpha+\gamma_1 \text{ and } \beta_1=\beta+\delta_1$$

and standard errors for your predictions. (These are the intercept and slope for the calibration equation for the 1st RTD.)

m) Now let's use the `R` mixed model routine to attempt some of what we've already done "by hand" with this calibration problem. With the `nlme` package added to your `R` session, make a file called (say) `thermocouples.txt` that looks like

```
y group temp
99.8 1 0
108.1 1 1
136.0 1 4
100.3 2 0
109.5 2 1
137.7 2 4
98.3 3 0
110.1 3 1
142.2 3 4
```

and read it into your `R` session as

```
thermo <- read.table("thermocouples.txt ", header=T)
```

Then prepare the data table for use in the fitting of a mixed model by typing

```
gd <- groupedData(y~temp|group,data=thermo)
```

Then issue commands that will do model fitting via maximum likelihood and REML

```
> fm1 <- lme(y~temp,data=gd,random=~1+temp|group,method="ML")
> fm2 <- lme(y~temp,data=gd,random=~1+temp|group,method="REML")
```

As it turns out, if you examine the fitted model objects `fm1` and `fm2` you will come to the conclusion that something other than what you've found to this point has been computed. This is because these calls allow the $(\gamma_i,\delta_i)'$ pairs to have a general covariance matrix, and the model we've been using specified that $\text{Cov}(\gamma_i,\delta_i)=0$. So we need to modify the fitting to agree with our modeling. Type

```
> fm3 <- update(fm1,random=pdDiag(~1+temp),method="ML")
> fm4 <- update(fm2,random=pdDiag(~1+temp),method="REML")
```

I believe that these calls create fitted model objects that correspond to the analysis we've done thus far. You may see some of what these calls have created by typing

```
> summary(fm3)
> summary(fm4)
```

But much more than this is available. Type

```
> ?nlmeObject
```

to see exactly what is available. Then apply all of the functions `random.effects()`, `fixed.effects()`, `predict()`,`coefficients()`, and `intervals()` to both fitted model objects. What parts of what these calls produce have you already calculated?

What do the results of the `intervals()` calls suggest about what this very small data set provides in terms of precision of estimation of the variance components? (You should have learned in a basic course that sample sizes required to really pin down variances or standard deviations are typically huge. Notice, for example, that we have only 3 RTDs … a sample of size 3 for estimating $\sigma_\gamma^2$ and $\sigma_\delta^2$.)

2. Do Part III of the Stat 511 question from the May 2003 Statistics MS Exam, posted on the current course Web page.

3. Do Question 2 of the Spring 2003 Stat 511 Exam II posted on the 2003 Stat 511 Web page.