This assignment consists of a number of problems on mixed linear models. Some of the problems are repeats of ones from HW 10 Spring 2003. You are responsible for the content of all the problems. However, only problems 2,3,4 and 6 should be turned in.

1. Below is a very small set of fake unbalanced 2 level nested data. Consider the analysis of these under the mixed effects model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$$

where the only fixed effect is $\mu$, the random effects are all independent with the $\alpha_i \ iid \ N\left(0,\sigma^2_\alpha\right)$, the $\beta_{ij} \ iid \ N\left(0,\sigma^2_\beta\right)$, and the $\varepsilon_{ijk} \ iid \ N\left(0,\sigma^2\right)$.

<table>
<thead>
<tr>
<th>Level of A</th>
<th>Level of B within A</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6.0, 6.1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>8.6, 7.1, 6.5, 7.4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9.4, 9.9</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>9.5, 7.5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6.4, 9.1, 8.7</td>
</tr>
</tbody>
</table>

Again after adding the MASS and nlme libraries to your R session (and specifying the use of the sum restrictions), do the following.

a) Enter your data by typing

```r
> y<-c(6.0,6.1,8.6,7.1,6.5,7.4,9.4,9.9,9.5,7.5,6.4,9.1,8.7)
> A<-c(1,1,1,1,1,1,2,2,2,2,2,2,2,2)
> B<-c(1,1,2,2,2,2,1,1,2,2,3,3,3,3)
```

Run and view some summaries of a random effects/mixed effects analysis of these data by typing

```r
> lme.out<-lme(y~1,random=~1|A/B)
> summary(lme.out)
```

Get approximate confidence intervals for the parameters of the mixed model by typing

```r
> intervals(lme.out)
```

Compute an exact confidence interval for $\sigma$ based on the mean square error (or pooled variance from the 5 samples of sizes 2,4,2,2, and 3). How do these limits compare to what R provides in this analysis?

b) A fixed effects analysis can be made here and some summaries viewed here by typing

```r
> AA<-as.factor(A)
> BB<-as.factor(B)
> lmf.out<-lm(y~1+AA/BB)
> summary(lmf.out)
```
(Note that the estimate of $\sigma$ produced by this analysis is exactly the one based on the mean square error referred to in a.) Type

```r
> predict(lmef.out2)
> predict(lmf.out)
```

Identify the predictions from the fixed effects model as simple functions of the data values. (What are these predictions?) Notice that the predictions from the mixed effects analysis are substantially different from those based on the fixed effects model.

2. The article “Variability of Sliver Weights at Different Carding Stages and a Suggested Sampling Plan for Jute Processing” by A. Lahiri (*Journal of the Textile Institute*, 1990) concerns the partitioning of variability in “sliver weight.” (A sliver is a continuous strand of loose, untwisted wool, cotton, etc., produced along the way to making yarn.) For a particular mill, 3 (of many) machines were studied, using 5 (10 mm) pieces of sliver cut from each of 5 rolls produced on the machines. The weights of the (75) pieces of sliver were determined and a standard hierarchical (balanced data) ANOVA table was produced as below. (The units of weight were not given in the original article.)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machines</td>
<td>1966</td>
<td>2</td>
</tr>
<tr>
<td>Rolls</td>
<td>644</td>
<td>12</td>
</tr>
<tr>
<td>Pieces</td>
<td>280</td>
<td>60</td>
</tr>
<tr>
<td>Total</td>
<td>2890</td>
<td>74</td>
</tr>
</tbody>
</table>

Use the same mixed effects model as in problem 1 and do the following.

a) Make 95% confidence intervals for each of the 3 standard deviations $\sigma_a, \sigma_\beta$ and $\sigma$. Based on these, where do you judge the largest part of variation in measured weight to come from? (From differences between pieces for a given roll? From differences between rolls from a given machine? From differences between machines?) (Use Cochran- Satterthwaite for $\alpha$ and $\beta$ and give an exact interval for $\sigma$.)

b) Suppose for sake of illustration that the grand average of all 75 weight measurements was in fact $\bar{y}_{..} = 35.0$. Use this and make a 95% confidence interval for the model parameter $\mu$.

3. According to a posting to an R-help list by Doug Bates “It happens that the `lme` function is much better suited to nested random effects than to crossed random effects. To estimate crossed random effects you must create an awkward formulation with a grouping factor that has one level and the random effects model matrix based on based on the indicators for factor a and for factor b. These two sets of random effects each have variance-covariance matrix that are multiples of an identity and are grouped together as a block-diagonal matrix.” (The context here is a 2-way factorial and a “no-interaction” analysis.)

Consider the fake unbalanced $3 \times 3$ factorial data of problem 4 on HW 5. Here let’s attempt a random effects analyses of those

$$y_{ijk} = k$$th response at level $i$ of A and level $j$ of B

based on a two-way random effects model without interaction
\[
y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}
\]
where \(\mu\) is an unknown constant, the \(\alpha_i\) are iid \(N(0, \sigma_{\alpha}^2)\), the \(\beta_j\) are iid \(N(0, \sigma_{\beta}^2)\), the \(\varepsilon_{ijk}\) are iid \(N(0, \sigma^2)\) and all the sets of random effects are all independent.

Load the \texttt{nlme} and \texttt{MASS} packages and then read the data into R and prepare a first appropriate “\texttt{groupedData}” object using the following commands.

```r
Y<-c(12,14,10,12,9,11,12,6,7,10,11,7)
a<-c(1,1,1,1,2,2,2,3,3,1,2,3)
b<-c(1,2,3,3,1,2,2,3,3,1,2,3)
group<-c(1,1,1,1,1,1,1,1,1,1,1,1)
A<-as.factor(a)
B<-as.factor(b)
Group<-as.factor(group)
Fake<-data.frame(Y,A,B,Group)
FakeGrouped<-groupedData(Y~1|Group,Fake)
```

Then the following command will run the mixed effects analysis on these data.

```r
> twoway.out<-lme(Y~1,data=FakeGrouped,random=pdBlocked(list(pdIdent(~A-1),pdIdent(~B-1))))
```

Run the \texttt{summary()}, \texttt{predict()}, and \texttt{intervals()} functions on the object \texttt{twoway.out}. What are 95% confidence limits for the 4 model parameters \(\mu, \sigma_{\alpha}, \sigma_{\beta}, \) and \(\sigma\)?

4. The data set below is taken from page 54 of \textit{Statistical Quality Assurance Methods for Engineers} by Vardeman and Jobe. It gives burst strength measurements made by 5 different technicians on small pieces of paper cut from 2 different large sheets (each technician tested 2 small pieces from both sheets).

<table>
<thead>
<tr>
<th>Technician (B)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheet (A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13.5, 14.8</td>
<td>10.5, 11.7</td>
<td>12.9, 12.0</td>
<td>8.8, 13.5</td>
<td>12.4, 16.0</td>
</tr>
<tr>
<td>2</td>
<td>11.3, 12.0</td>
<td>14.0, 12.5</td>
<td>13.0, 13.1</td>
<td>12.6, 12.7</td>
<td>11.0, 10.6</td>
</tr>
</tbody>
</table>

Vardeman and Jobe present an ANOVA table for these data based on a two-way random effects model \textit{with} interaction

\[
y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}
\]

(where \(\mu\) is the only fixed effect).

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheet (A)</td>
<td>.5445</td>
<td>1</td>
<td>10(\sigma_{\alpha}^2 + 2\sigma_{\alpha\beta}^2 + \sigma^2)</td>
</tr>
<tr>
<td>Technician (B)</td>
<td>2.692</td>
<td>4</td>
<td>4(\sigma_{\beta}^2 + 2\sigma_{\alpha\beta}^2 + \sigma^2)</td>
</tr>
<tr>
<td>Sheet<em>Technician (A</em>B)</td>
<td>24.498</td>
<td>4</td>
<td>2(\sigma_{\alpha\beta}^2 + \sigma^2)</td>
</tr>
<tr>
<td>Error</td>
<td>20.955</td>
<td>10</td>
<td>(\sigma^2)</td>
</tr>
<tr>
<td>Total</td>
<td>48.6895</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>
a) Use R and an ordinary linear model analysis to verify that the sums of squares above are correct.

b) In the context of a “gauge study” like this, \( \sigma \) is usually called the “repeatability” standard deviation. Give a 95% exact confidence interval for \( \sigma \) (based on the error mean square in the ANOVA table.)

c) Under the two-way random effects model above, the quantity \( \sqrt{\sigma^2_\beta + \sigma^2_{a\beta}} \) is the long run standard deviation that would be seen in many technicians measure from the same sheet in the absence of any repeatability variation. It is called the “reproducibility” standard deviation. Find a linear combination of mean squares that has expected value the reproducibility variance, use Cochran-Satterthwaite and make approximate 95% confidence limits for that variance, and then convert those limits to limits for the reproducibility standard deviation here.

d) Now let’s use lme in an analysis of these data. Enter the data, set up an appropriate groupedData object and run the lme function using the commands below.

```r
Y<- c(13.5,14.8,10.5,11.7,12.9,12.0,8.8,13.5,12.4,16.0,11.3,12.0,14.0, 12.5,13.0,13.1,12.6,12.7,11.0,10.6) > a<-c(rep(1,10),rep(2,10)) > b<-c(1,1,2,2,3,3,4,4,5,5,1,1,2,2,3,3,4,4,5,5) > ab<-c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10) > group<-c(rep(1,20)) > A<-as.factor(a) > B<-as.factor(b) > AB<-as.factor(ab) > Group<-as.factor(group) > Gauge<-data.frame(Y,A,B,AB,Group) > GaugeGrouped<-groupedData(Y~1|Group,Gauge) > gauge.out<-lme(Y~1,data=GaugeGrouped,random=pdBlocked(list(pdIdent(~A-1),pdIdent(~B-1),pdIdent(~AB-1))))
```

Run the summary(), predict(), and intervals() functions on the object gauge.out. What are point estimates of the repeatability and reproducibility standard deviation based on REML? In light of the confidence limits for the standard deviations that R prints out, how reliable do you think the point estimate of reproducibility standard deviation is? Does your approximate interval from c) really do justice to the uncertainty here? (In retrospect, it should be obvious that one can’t learn much about technician-to-technician variability based on a sample of 5 technicians!)
5. Below is a small set of fake unbalanced 2-way factorial (in factors A and B) data from 2 (unbalanced) blocks.

<table>
<thead>
<tr>
<th>Level of A</th>
<th>Level of B</th>
<th>Block</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9.5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>11.3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>8.4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>8.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6.7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9.4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>6.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12.2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>13.1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>10.7</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>8.9</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10.8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Consider an analysis of these data based on the model

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \tau_k + \varepsilon_{ijk} \]  

(*)

first where (*) specifies a fixed effects model (say under the sum restrictions) and then where the \( \tau \) are iid \( N(0, \sigma^2_\tau) \) random block effects independent of the iid \( N(0, \sigma^2) \) random errors \( \varepsilon_{ijk} \).

Begin by adding the MASS and nlme libraries to your R session.

a) Create vectors \( A, B, Block, y \) in R, corresponding to the columns of the data table above. Make from the first 3 of these objects that R can recognize as factors by typing

\[
\begin{align*}
> & A<-\text{as.factor}(A) \\
> & B<-\text{as.factor}(B) \\
> & FBlock<-\text{as.factor}(Block)
\end{align*}
\]

Tell R you wish to use the sum restrictions in linear model computations by typing

\[
\begin{align*}
> & \text{options(contrasts=c("contr.sum", "contr.sum"))}
\end{align*}
\]

Then note that a fixed effects linear model with equation (*) can be fit and some summary statistics viewed in R by typing

\[
\begin{align*}
> & \text{lmf.out<-lm(y~A*B+FBlock)} \\
> & \text{summary(lmf.out)}
\end{align*}
\]

What is an estimate of \( \sigma \) in this model?

b) The linear mixed effects model (*) may be fit and some summaries viewed by typing

\[
\begin{align*}
> & \text{lmef.out<-lme(y~A*B, random=~1|Block)}
\end{align*}
\]
In terms of inferences for the fixed effects, how much difference do you see between the outputs for the analysis with fixed block effects and the analysis with random block effects?

How do the estimates of $\sigma$ compare for the two models? Type

```
> intervals(lmef.out)
```

to get confidence intervals for the mixed effects model. Compare a 95% confidence interval for $\sigma$ in the fixed effects model to what you get from this call.

In the fixed effects analysis of a), one get estimates of $\tau_2 = -\tau_1$. Are these consistent with the estimate of $\tau_1^2$ produced here? Explain.

c) Type

```
> coef(lmf.out)
> coef(lmef.out)
> fitted(lmf.out)
> fitted(lmef.out)
```

What are the BLUE and (estimated) BLUP of $\mu + \alpha_i + \beta_j + \alpha \beta_{ij} + \tau_i$ in the two analyses? Notice that in the random blocks analysis, it makes good sense to predict a new observation from levels 1 of A and 1 of B, from an as yet unseen third block. This would be $\mu + \alpha_i + \beta_j + \alpha \beta_{ij} + \tau_3$. What value would you use to predict this? Why does it not make sense to do predict a new observation with mean $\mu + \alpha_i + \beta_j + \alpha \beta_{ij} + \tau_3$ in the fixed effects context?

6. Consider the scenario of the example beginning on page 1190 of Neter, Wasserman, and friends. This supposedly concerns sales of pairs of athletic shoes under two different advertising campaigns in several test markets. The data from Neter, et al. Table 29.10 are on page 8 of this assignment.

Enter the data into R as 5 vectors of length 30 called `time`, `ad`, `test`, `loc`, `y`. Notice that each combination of ad campaign level and an index value for test market for that campaign is associated with a different level of the locality variable. (Making the locality vector is the only device Vardeman was able to come up with to get the right mixed effects model run.) We consider an analysis of these data based on the “split plot”/“repeated measures” mixed effects model

$$
y_{ijk} = \mu + \alpha_i + \beta_j + \alpha \beta_{ij} + \gamma_{jk} + \epsilon_{ijk}
$$

for $y_{ijk}$ the response at time $i$ from the $k$th test market under ad campaign $j$, where the random effects are the $\gamma_{jk}$'s and the $\epsilon_{ijk}$'s.

Again after adding the MASS and nlme libraries to your R session (and specifying the use of the sum restrictions), do the following:

Type

```
> TIME<-as.factor(time)
> AD<-as.factor(ad)
```
> TEST<-as.factor(test)
> LOC<-as.factor(loc)

a) Neter, Wasserman, and friends produce a special ANOVA table for these data. An ANOVA table with the same sums of squares and degrees of freedom can be obtained in R using fixed effects calculations by typing

```r
> lmf1.out<-lm(y~1+TIME*(AD/TEST))
> anova(lmf1.out)
```

Identify the sums of squares and degrees of freedom in this table by the names used in class (SSA, SSB, SSAB, SSC, SSE, etc.). Then run a fictitious 3-way full factorial analysis of these data by typing

```r
> lmf2.out<-lm(y~1+TIME*AD*TEST)
> anova(lmf2.out)
```

Verify that the sums of squares and degrees of freedom in the first ANOVA table can be gotten from this second table. (Show how.)

b) A mixed model analysis in R can be gotten by typing

```r
> lmef.out<-lme(y~1+TIME*AD,random=~1|LOC)
> summary(lmef.out)
> intervals(lmef.out)
> predict(lmef.out)
```

What are 95% confidence intervals for $\gamma$ and $\sigma$ read from this analysis?

c) Make exact 95% confidence limits for $\sigma$ based on the chi-square distribution related to $MSE$. Then use the Cochran-Satterthwaite approximation to make 95% limits for $\sigma_\gamma$. How do these two sets of limits compare to what R produced in part b)?

d) Make 95% confidence limits for the difference in Time 1 and Time 2 main effects. Make 95% confidence limits for the difference in Ad Campaign 1 and Ad Campaign 2 main effects. How can an interval for this second quantity be obtained nearly directly from the `intervals(lmef.out)` call made above?

e) How would you predict the Time 2 sales result for Ad Campaign 2 in another (untested) market? (What value would you use?)
Data from Neter, et al. Table 29.10

<table>
<thead>
<tr>
<th>Time Period A</th>
<th>Ad Campaign B</th>
<th>Test Market C(B)</th>
<th>Locality</th>
<th>Response (Sales?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>958</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1,005</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>351</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>549</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>730</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>780</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>229</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>883</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>624</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>375</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1,047</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1,122</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>436</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>632</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>784</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>897</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>275</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>964</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>695</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>436</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>933</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>986</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>339</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>512</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>707</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>718</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>202</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>817</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>599</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>351</td>
</tr>
</tbody>
</table>