Homework 3

Each part of the problems 5 points
Due in class or by email before 9am on Thursday February 10

1. KNNL 25.17
(Note: you can choose either the restricted or the unrestricted version of the model.
Please state clearly which model you use.)

2. KNNL 25.27 (a)-(d)

3. [Methods Qualifying Exam, January 2010: use paper and pencil.]
A 3 x 4 factorial study was designed to investigate the effects of 3 fertilization methods (factor A) and 4 seeding rates (factor B) on the yield of sugar beets. To this end they conducted an experiment with the combinations of all levels of both factors. Each treatment combination was replicated in three plots. The two tables below show the mean yield values $\bar{y}_{ij}$ for each treatment combination, as well as some summaries.

<table>
<thead>
<tr>
<th>Level of A</th>
<th>Level of B</th>
<th>$\bar{y}_{ij}$</th>
<th>Source</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>15.87</td>
<td>16.3</td>
<td>17.19</td>
<td>17.66</td>
</tr>
<tr>
<td>2</td>
<td>17.41</td>
<td>17.63</td>
<td>18.12</td>
<td>18.51</td>
</tr>
<tr>
<td>3</td>
<td>14.81</td>
<td>15.73</td>
<td>16.32</td>
<td>16.8</td>
</tr>
</tbody>
</table>

(a) i. State the ANOVA model, and the corresponding assumptions, that can be used to answer research questions in this experiments.
ii. Test whether the change in yield due to the use of different fertilizers depends on the seedling rate. State the null and the alternative hypotheses, and the conclusions. Use the significance level of 5%.
iii. Test whether there is a difference in yield between the fertilizers. State the null and the alternative hypotheses, and the conclusions. Use the significance level of 5%.

(b) Test whether the use of fertilizers 1 and 2 results in a same mean yield. State the null and the alternative hypotheses, and the conclusions. Use the significance level of 5%.

(c) Suppose now that instead of focusing on 4 levels of seeding rate, the researchers were interested in variations in yield that can result from varying the seeding rate over the entire operational range. Therefore, the four seeding rates above are not the fixed levels of interest, but a random sample of all possible levels.
   i. Modify the model above to reflect the different experimental setting, and state the assumptions.
ii. Using the modified model, test whether the change in yield due to the use of different fertilizers depends on the seedling rate. State the null and the alternative hypotheses, and the conclusions. Use the significance level of 5%.

iii. Using the modified model, test whether there is a difference in yield between the fertilizers. State the null and the alternative hypotheses, and the conclusions. Use the significance level of 5%.

iv. Test whether the use of fertilizers 1 and 2 results in a same mean yield, given this new model specification. State the null and the alternative hypotheses, and the conclusions. Use the significance level of 5%. Compare with the results of (b).

4. [Methods Qualifying Exam, January 2005: use paper and pencil.]
The school superintendent is concerned about the development of technology skills in middle school. Since there are 3 middle schools in his district, all of which go about this instruction differently, he decided to assess if there were any differences across schools. He first compiled a long list of “tech skills” and randomly selected 5 to be used in his study. He then randomly selected 20 students from each school and assigned each to one of the five tasks so that there were 4 students per task per school. Each student then performed the skill and a score between 0 and 100 was assigned.

(a) If a two-way ANOVA is to be used for the analysis, should it be treated as a fixed effects, random effects, or mixed effects model? Explain.

(b) Complete the following ANOVA table and determine which effects are significant at the $\alpha = .05$ level. State your conclusions, making sure to estimate all variances and describing any additional mean comparisons you’d like to perform.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td></td>
<td>220.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill</td>
<td></td>
<td>96.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School $\times$ Skill</td>
<td></td>
<td>176.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>450.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) If the grand skill level of the middle schools (average over the three schools) is of interest, describe how one would construct a 95% confidence interval.

5. [Methods Qualifying Exam, January 2009: use paper and pencil.]
The following dataset concerns an experiment where six pullets were placed into each of 12 pens. Four blocks were formed from groups of three pens based on location. Three treatments were applied. The number of eggs produced was recorded.

<table>
<thead>
<tr>
<th>treat block eggs</th>
<th>treat</th>
<th>block</th>
<th>eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 1 330</td>
<td>1</td>
<td>0</td>
<td>330</td>
</tr>
<tr>
<td>2 0 2 288</td>
<td>2</td>
<td>0</td>
<td>288</td>
</tr>
<tr>
<td>3 0 3 295</td>
<td>3</td>
<td>0</td>
<td>295</td>
</tr>
<tr>
<td>4 0 4 313</td>
<td>4</td>
<td>0</td>
<td>313</td>
</tr>
<tr>
<td>5 E 1 372</td>
<td>5</td>
<td>E</td>
<td>372</td>
</tr>
</tbody>
</table>
An analysis on this dataset using lme in nlme package provides the output (Note: lme is an older implementation of mixed effects models in R, and has a slightly different model syntax. If the syntax presents challenges, you can refit the model with lmer using the data above.)

```r
> remlme1 <- lme(eggs~treat,random=~1|block,data=eggprod,method=REML)
> summary(remlme1)
Linear mixed-effects model fit by REML
Data: eggprod
AIC BIC logLik
95.41439 96.40051 -42.70719

Random effects:
Formula: ~1 | block
(Intercept) Residual
StdDev: 11.39932 19.67020

Fixed effects: eggs ~ treat
Value Std.Error DF  t-value p-value
(Intercept) 349.00 11.36729 6 30.702129 0.0000
  treatF -6.25 13.90893 6  -0.449352 0.6690
  treatO -42.50 13.90893 6  -3.055591 0.0224
```

(a) Describe the model used in the above analysis (in terms of \( Y = X\beta + Z\gamma + \varepsilon \), where \( X \) includes all covariates with fixed effects and \( Z \) includes all covariates with random effects), and report the estimates of all parameters describing the model.

(b) Give reason to use REML instead of ML to fit the model.

(c) A new pen is added to the first block and treated with treatment F. Please predict the number of eggs for this pen along with an estimate of the variability in this prediction. What if this new pen is placed in a new location (i.e., a fifth block)?

```r
> fixed.effects(remlme1)
(Intercept) treatF treatO
  349.00  -6.25  -42.50
> random.effects(remlme1)
(Intercept)
1 10.497605  2  -5.562476  3  2.132979  4  -7.068108
```
(d) We also tried other methods and other models on this dataset, and got the log-likelihood function values:

```r
> remlme2 <- lme(eggs~1, random=~1|block, data=eggprod, method=REML)
> logLik(remlme2)
log Lik. -53.65615
> mlme1 <- lme(eggs~treat, random=~1|block, data=eggprod, method=ML)
> logLik(mlme1)
log Lik. -52.44424
> mlme2 <- lme(eggs~1, random=~1|block, data=eggprod, method=ML)
> logLik(mlme2)
log Lik. -56.65651
```

(e) Please determine the significance of the treatment using a likelihood ratio test (assuming $\chi^2$ distribution). Should you improve the accuracy of this p-value? If yes, please state your approach.

(f) Please propose an approach to testing whether there is a significant difference between the blocks, and state how you would like to calculate your test statistics and the p-value.