Follow-up on corn yield analysis

The artificial corn yield data had the following questions.

1. Is high or low irrigation better for larger crop yield? Does the answer depend on the variety? Estimate the size of important effects and provide error estimates.

2. Which variety or varieties yield the most? Does the answer depend on the irrigation level? Estimate the size of important effects and provide error estimates.

A graphical analysis (and a formal analysis, if you did one), showed that an interaction term should be included in the model. So, when examining irrigation, variety matters and vice versa. The subsequent question on how to account for field is best answered as a random effect — field is a nuisance variable with many (eight, here) different values, and one can imagine the fields selected for the experiment as a subset of the larger collection of fields to which we would hope that the results of the analysis would apply. Consequently, a reasonable model that many of you selected is, in R package `lmer` terms, the following.

```
corn = read.table("yield.txt", header=T)
corn$Field = factor(paste("Field",corn$Field))
require(lme4)
corn.lmer = lmer(Yield ~ Variety*Irrigation + (1 | Field), data=corn)
```

Please note that the previous model was fit using the default REML method, which is beneficial for parameter estimation, but is problematic for hypothesis testing. We will get into this issue later in the semester.

Now, what is meant by the word *effect* and how does this depend on the interaction between irrigation and variety? I would say for this problem that the effect of irrigation needs to be quantified separately for each variety and is best expressed as the estimated difference between yields at high and low irrigation levels for each variety. Using notation that I hope is self-explanatory, I want (and think the client would want) to see estimates (with measures of uncertainty) of these quantities.

$$\mu_{A,high} - \mu_{A,low} \qquad \mu_{B,high} - \mu_{B,low} \qquad \mu_{C,high} - \mu_{C,low} \qquad \mu_{D,high} - \mu_{D,low}$$

Of course, switching high and low is equivalent. The summary of `corn.lmer` will have estimates and p-values for eight different fixed parameters, but not all of these correspond to questions of primary interest. How do you estimate these effects and find confidence intervals for them?

Similarly, when seeking the effect of variety, irrigation is going to matter. One could potentially ask to make all possible pairwise comparisons between two varieties at each irrigation level. Another option is to pick one variety at each irrigation level (say the observed best) and estimate how much worse each of the others is in comparison. At high irrigation, variety B appears to the be best. What is the expected yield for variety B at high irrigation and how much worse are the other varieties? What is the expected yield for variety A at low irrigation and how much worse are the other varieties?

Think about how to do these estimates and confidence intervals before class on Tuesday. You may need to do more than pull information from this summary.

```
summary(corn.lmer)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Yield ~ Variety * Irrigation + (1 | Field)
##    Data: corn
##
## REML criterion at convergence: 100.6
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Field    (Intercept) 12.899   3.591
##  Residual              0.875   0.935
## Number of obs: 32, groups: Field, 8
##
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)              116.825      1.856    63.0
## VarietyB                   8.550      0.661    12.9
## VarietyC                   6.725      0.661    10.2
## VarietyD                   2.050      0.661     3.1
## Irrigationlow              6.750      2.624     2.6
## VarietyB:Irrigationlow  -15.225      0.935   -16.3
## VarietyC:Irrigationlow  -14.100      0.935   -15.1
## VarietyD:Irrigationlow   -5.400      0.935    -5.8
##
## Correlation of Fixed Effects:
##             (Intr) VartyB VartyC VartyD Irrgtn VrtB:I VrtC:I
## VarietyB    -0.178
## VarietyC    -0.178  0.500
## VarietyD    -0.178  0.500  0.500
## Irrigatinlw -0.707  0.126  0.126  0.126
## VrtyB:Irrgt  0.126 -0.707 -0.354 -0.354 -0.178
## VrtyC:Irrgt  0.126 -0.354 -0.707 -0.354 -0.178  0.500
## VrtyD:Irrgt  0.126 -0.354 -0.354 -0.707 -0.178  0.500  0.500
```