We are looking at a study that is examining the relationship between soybean growth and leaflet number. We wish to compare the mean leaf area for two varieties (3-leaflets and 7-leaflets lines) and for four different genotypes of soybean (Hendrick, Mn1401, MN1801, Traill).

The design of the field experiment is to split the farm up into 4 blocks. Each block will be partitioned into 8 sub-plots, to which are randomly allocated 1 replication of the two treatment factors in a crossed design. We are specifically interested in choosing between these two varieties (the leaflet numbers) and the 4 genotypes. We would like to extend our results to other plots in the field that aren't directly a part of the experiment.

1. Should genotype be considered a fixed or random effect? Why?
2. Should leaflet number be considered a fixed or random effect? Why?
3. Should block be considered a fixed or random effect? Why?

The average leaflet area for each plant is the response, *Y* (so over 3 leaflets or 7 leaflets, depending on the variety). We are interested in fitting the following model (here, I'm using a representative letter for notation instead "μ_j" to make the model easier to read and to not answer the previous questions with the notation)

$$Y_{ijkl} = \mu + G_j + L_k + GL_{jk} + B_l + \varepsilon_{ijkl}$$

Where

- G_j is the genotype ($j = 1,2,3,4$)
- L_k is the leaflet number ($k = 1,2$)
- B_l is the block ($l = 1,2,3,4$)

Additionally, specify all of the random effects and ε_{ijkl} as random, independent normal.

(you can use the generic $\mathbb{X}$ and $\mathbb{Z}$ notation, but you should indicate what the first observation's entry (that is, the first row) in those matrices would look like in both the conditional and marginal models)

4. What is the conditional model?

Conditional Mean:

Conditional Covariance:

Conditional Distribution:

5. What is the marginal model?

Marginal Mean:

Marginal Covariance:

Marginal Distribution:

I'm going to write down partial SAS code for doing the analysis. You should complete it to answer

the remaining questions:

```
/* Code for students to complete */
TITLE 'conditional model: ';
PROC GLIMMIX DATA = soy NOBOUND PLOTS = STUDENTPANEL(blup);
   CLASS <?>;
   MODEL Area = <?> / DDFM=KENWARDROGER;
   RANDOM <?>;
   LSMEANS geno*leaflet /  PLOT = meanplot(sliceby=leaflet join cl) DIFF;
RUN;
```

Some comments on this code:

- NOBOUND instructs SAS to allow a negative variance estimate.
- The STUDENTPANEL(blup) forms the diagnostic plots for checking assumptions (you do plan to check the assumptions, don't you?). blup is an acronym for "best linear unbiased predictor". Recall that in order to get residuals, we need to get fitted values. In order to get fitted values, we need to predict the values for the random effect(s). We are using something called the blup to do this.
- We are considering the interaction model as we don't have a good (scientific or statistical) reason to not include it.
- The SLICEBY command tells SAS to include the profile plots with leaflet being the two "profiles"

6. Write up a statistical conclusion addressing the assumptions and conclusions of this study. Are the random effects assumptions met? Is it possible to check them? Which combination of leaflet/genotype would you recommend?

7. Write up the code for fitting the marginal model. How do the results compare to the results from the conditional model?

8. Lastly, refit the conditional model without the "NOBOUND" option. How do the results compare to the output for the marginal model now?