HW 6

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?

Genotype is a fixed effect because its levels are prespecified rather than randomly sampled, and its effect is of research interest.

1. Should leaflet number be considered a fixed or random effect? Why?

Leaflet number is a fixed effect because its levels are prespecified rather than randomly sampled, and its effect is of research interest.

1. Should block be considered a fixed or random effect? Why?

Block is a random effect, because its effect is considered being sampled from a broader population of varying farm fields (effect of farm varies randomly from farm to farm), and its effect is not of research interest.

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 “to make the model easier to read and to not answer the previous questions with the notation)

Where

* is the genotype (
* is the leaflet number (
* is the block (

Additionally, specify all of the random effects and as random, independent normal.

(you can use the generic and 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)

1. What is the conditional model?

Conditional Mean:

Conditional Covariance:

Conditional Distribution:

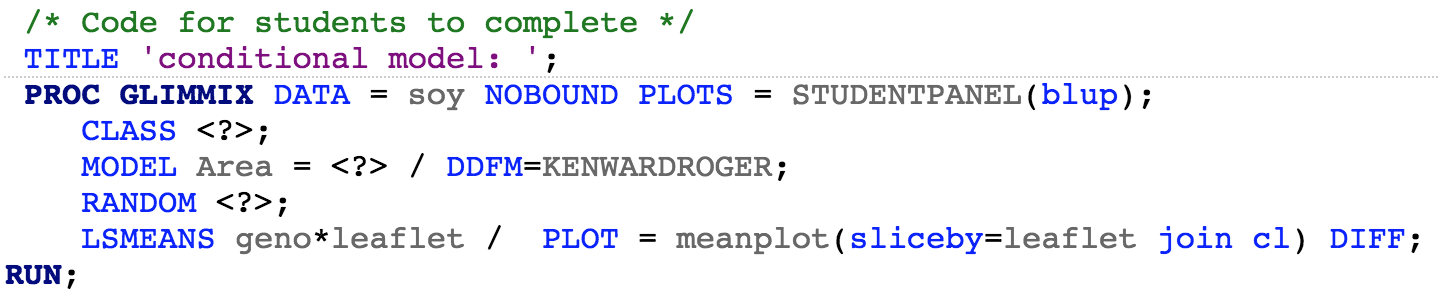
1. What is the marginal model?

Marginal Mean:

Marginal Covariance:

Marginal Distribution:

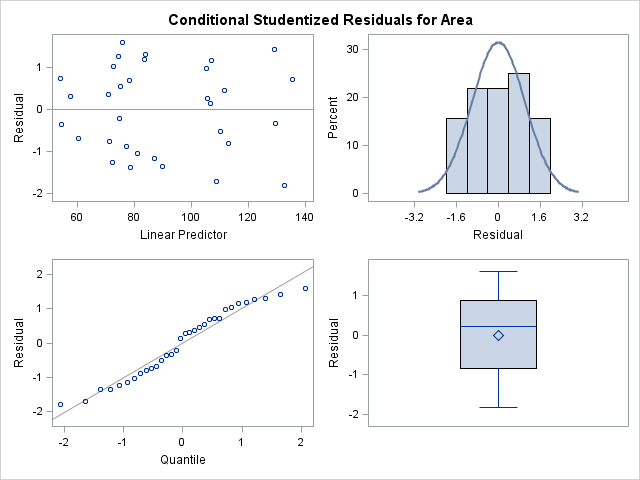
I’m going to write down partial SAS code for doing the analysis. You should complete it to answer the remaining questions:

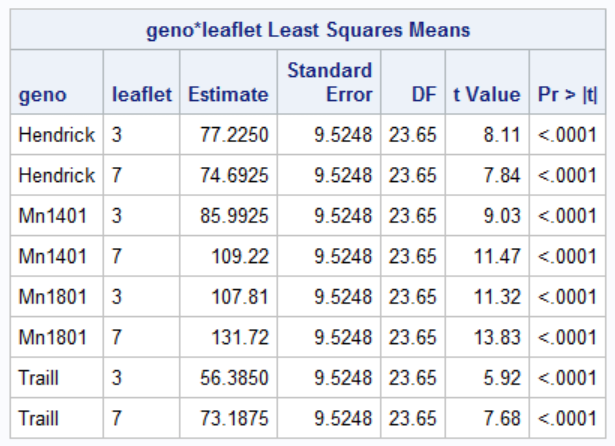
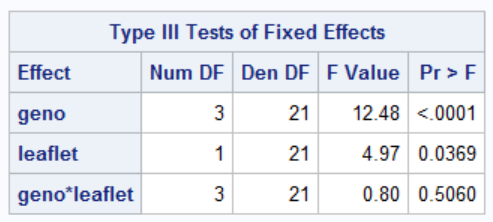


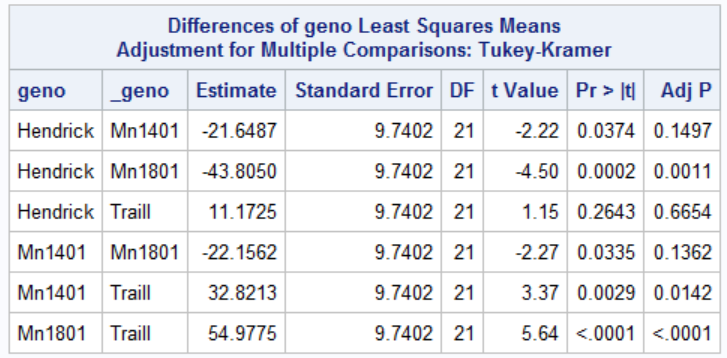
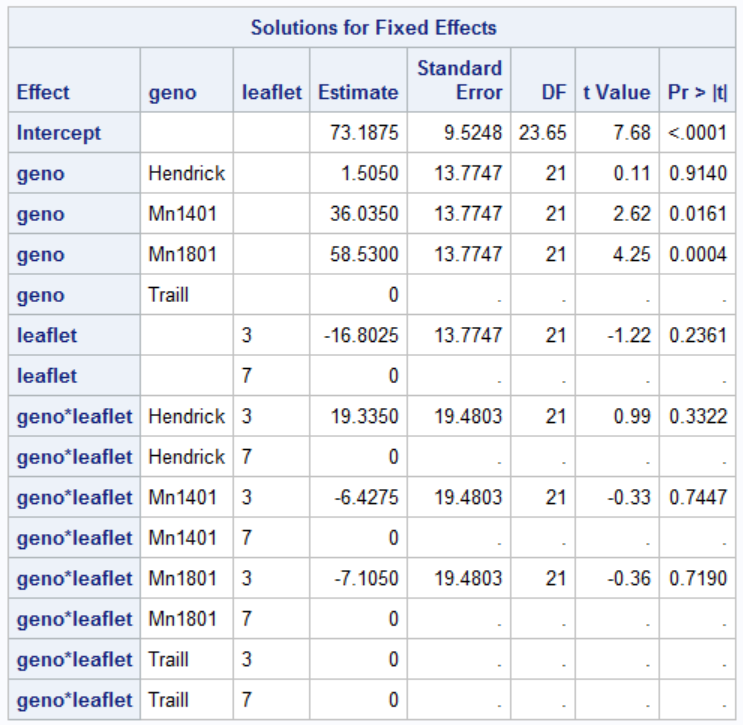
Some comments on this code:

* 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”

1. 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?







**Conclusion:**

Fitting a conditional mean model, the residual plots show slightly increasing variance, but normality assumption about residuals appear to be reasonable. There is little evidence (p-value = 0.5060) for interaction effect of genotype and number of leaflets on mean leaflet area, but strong evidence for main effects (p-value < 0.0001 for genotype, and p-value = 0.0369 for number of leaflets). Genotype Mn1081 has largest incremental effect on mean leaflet area, followed by Mn1401, and the two are not significantly different after Tuckey’s adjustment. Plants with 7 leaflets have higher mean leaflet area. If higher leaflet area is desired , genotype Mn1801 and 7 leaflet is recommended, followed by Mn1401 leaflet 7.

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

Residual plots from fitting a marginal model do not reveal violation of assumptions in terms of constant variance. The residuals appear to be not normal but still symmetric.

Parameter estimates and least square means are the same as fitting conditional model, so inferences would remain the same.

TITLE 'marginal model: ';

**PROC** **GLIMMIX** DATA = soy NOBOUND PLOTS = STUDENTPANEL(blup) outdesign=matrix;

CLASS block geno leaflet;

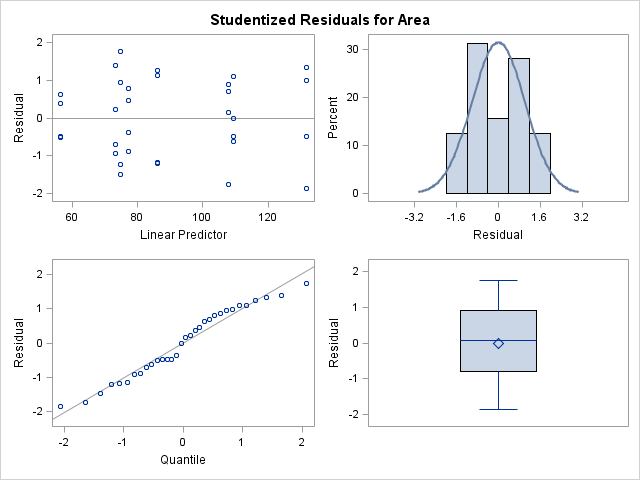
MODEL Area = geno leaflet geno\*leaflet /solution ddfm=kr;

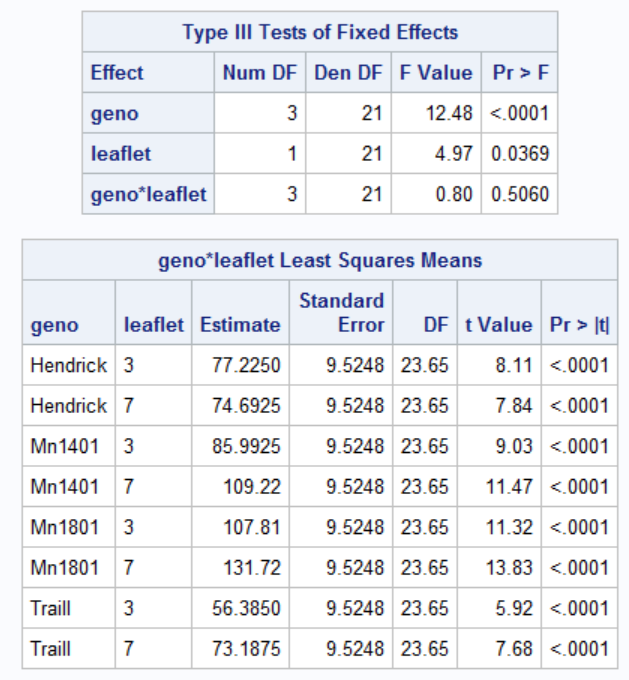
RANDOM \_residual\_ / solution TYPE = cs SUBJECT = block V;

LSMEANS geno\*leaflet / PLOT = meanplot(sliceby=leaflet join cl);

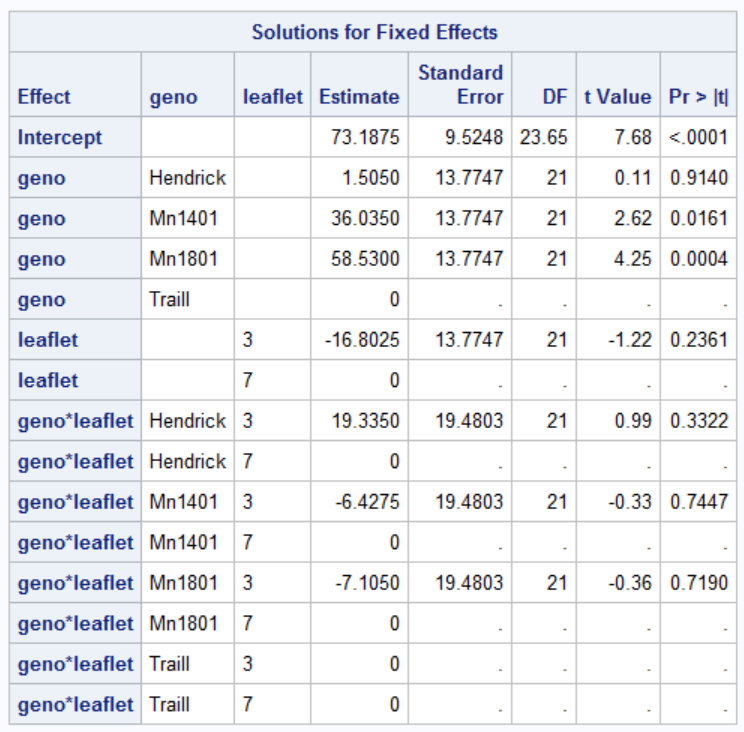
LSMEANS geno / adjust = tukey;

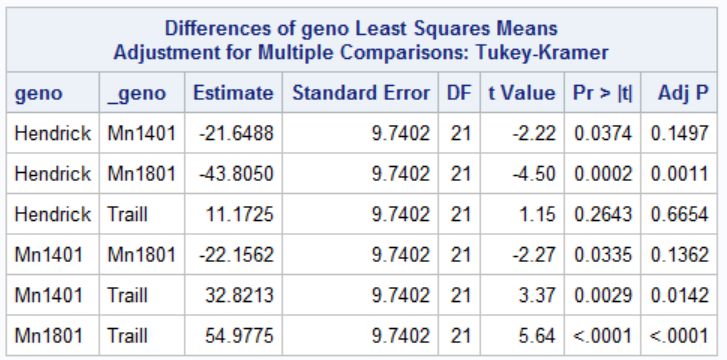
**RUN**;

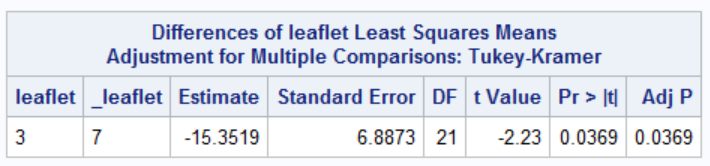




4







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

Without “NOBOUND” option, SAS would not allow estimate of to be negative in the conditional model, forcing the covariance matrix of Y to be . The marginal model would not have this problem, and estimates from it remains valid.

Conditional Model Marginal Model

