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.
2. 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.
3. 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 "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: $\mu\{Y_{ijkl}|B_l\} = \mu + G_j + L_k + GL_{jk} + B_l$

Conditional Covariance: $\sigma^2\boldsymbol{I_{nxn}}$

Conditional Distribution: $Y_{ijkl}|B_l \sim N(\mu + G_j + L_k + GL_{jk} + B_l, \sigma^2\boldsymbol{I_{nxn}})$

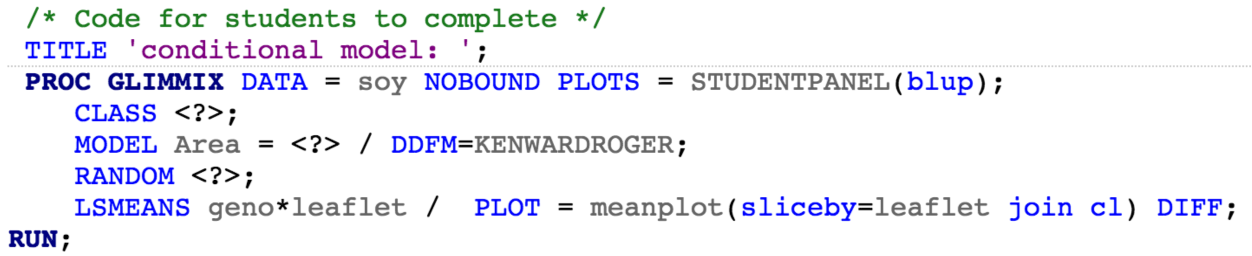
5. What is the marginal model?

Marginal Mean: $\mu\{Y\} = \mu + G_j + L_k + GL_{jk}$

Marginal Covariance: $\boldsymbol{\Sigma} = Cov\left(Y_{jkl}, Y_{j'k'l'}\right) = \begin{cases} \sigma^2 + \sigma_B^2 & if \quad j = j', \;\; k = k', l = l' \\ \sigma_B^2 & if \quad \left(j \neq j' \; or \; k \neq k'\right), and \; l = l' \\ 0 & otherwise \end{cases}$

Marginal Distribution: $Y_{ijkl}| \sim N(\mu + G_j + L_k + GL_{jk}, \boldsymbol{\Sigma})$

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

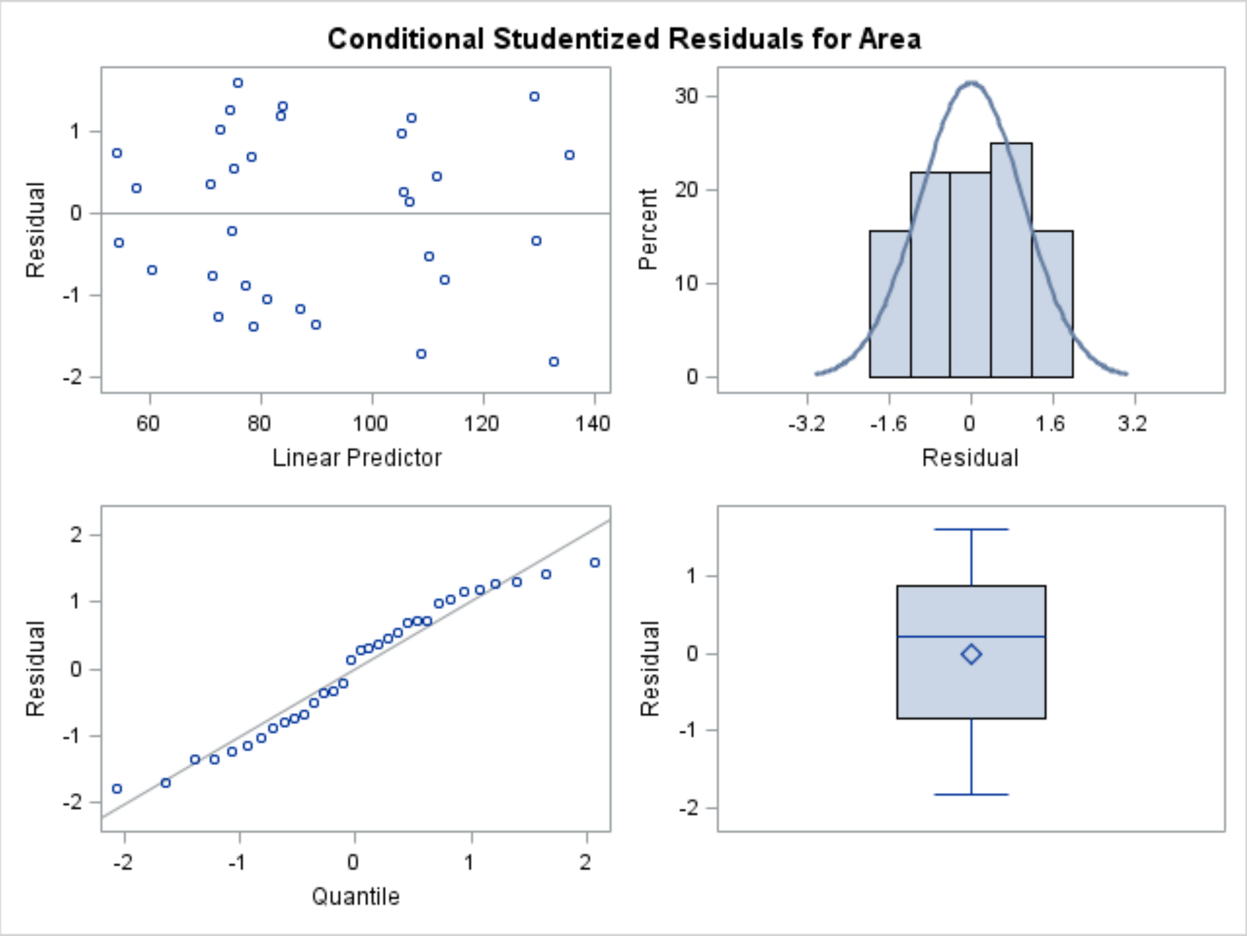


```
 /* Code for students to complete */
 TITLE 'conditional model: ';
 PROC GLIMMIX DATA = soy NOBOUND PLOTS = STUDENTPANEL(blup);
    CLASS <?>;
    MODEL Area = <?> / DDFM=KENWARDROGER;
    RANDOM <?>;
    LSMEANS geno*leaflet /  PLOT = meanplot(sliceby=leaflet join cl) DIFF;
RUN;
```

Some comments on this code:

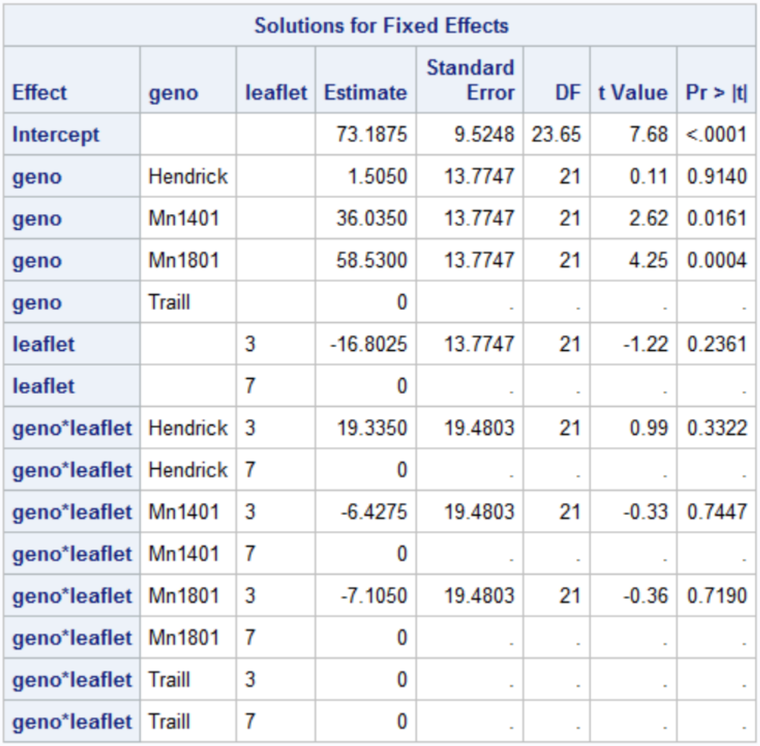
- NOBOUND instructs SAS to allow a negative variance estimate.
- The STUDENTPANEL(blup) forms the diagnostic plots for checking assumptions (you do plan to check the assumptions, don't you?). blup is an acronym for "best linear unbiased predictor". Recall that in order to get residuals, we need to get 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?



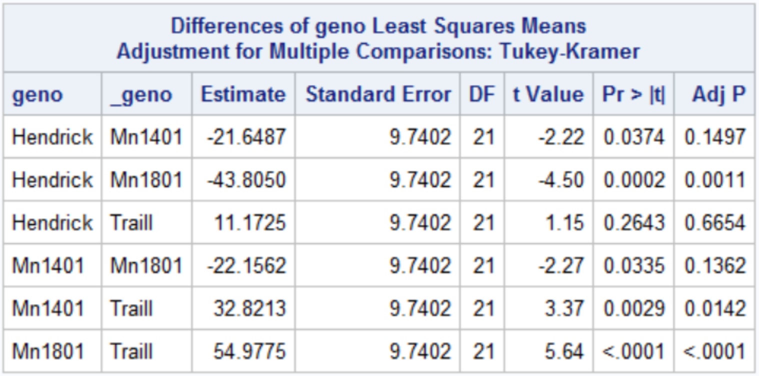
geno*leaflet Least Squares Means						
geno	leaflet	Estimate	Standard Error	DF	t Value	Pr > |t|
Hendrick	3	77.2250	9.5248	23.65	8.11	<.0001
Hendrick	7	74.6925	9.5248	23.65	7.84	<.0001
Mn1401	3	85.9925	9.5248	23.65	9.03	<.0001
Mn1401	7	109.22	9.5248	23.65	11.47	<.0001
Mn1801	3	107.81	9.5248	23.65	11.32	<.0001
Mn1801	7	131.72	9.5248	23.65	13.83	<.0001
Traill	3	56.3850	9.5248	23.65	5.92	<.0001
Traill	7	73.1875	9.5248	23.65	7.68	<.0001

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
geno	3	21	12.48	<.0001
leaflet	1	21	4.97	0.0369
geno*leaflet	3	21	0.80	0.5060



Solutions for Fixed Effects

Effect	geno	leaflet	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			73.1875	9.5248	23.65	7.68	<.0001
geno	Hendrick		1.5050	13.7747	21	0.11	0.9140
geno	Mn1401		36.0350	13.7747	21	2.62	0.0161
geno	Mn1801		58.5300	13.7747	21	4.25	0.0004
geno	Traill		0	.	.	.	.
leaflet		3	-16.8025	13.7747	21	-1.22	0.2361
leaflet		7	0	.	.	.	.
geno*leaflet	Hendrick	3	19.3350	19.4803	21	0.99	0.3322
geno*leaflet	Hendrick	7	0	.	.	.	.
geno*leaflet	Mn1401	3	-6.4275	19.4803	21	-0.33	0.7447
geno*leaflet	Mn1401	7	0	.	.	.	.
geno*leaflet	Mn1801	3	-7.1050	19.4803	21	-0.36	0.7190
geno*leaflet	Mn1801	7	0	.	.	.	.
geno*leaflet	Traill	3	0	.	.	.	.
geno*leaflet	Traill	7	0	.	.	.	.



Differences of geno Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

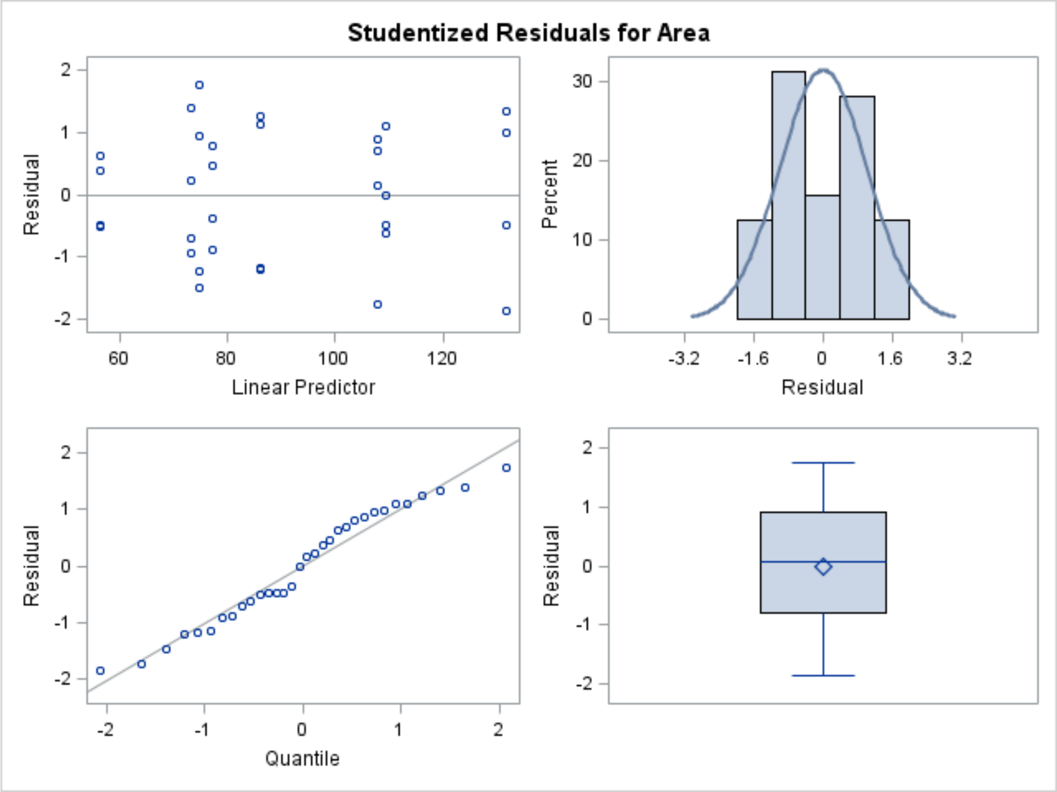
geno	_geno	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Hendrick	Mn1401	-21.6487	9.7402	21	-2.22	0.0374	0.1497
Hendrick	Mn1801	-43.8050	9.7402	21	-4.50	0.0002	0.0011
Hendrick	Traill	11.1725	9.7402	21	1.15	0.2643	0.6654
Mn1401	Mn1801	-22.1562	9.7402	21	-2.27	0.0335	0.1362
Mn1401	Traill	32.8213	9.7402	21	3.37	0.0029	0.0142
Mn1801	Traill	54.9775	9.7402	21	5.64	<.0001	<.0001

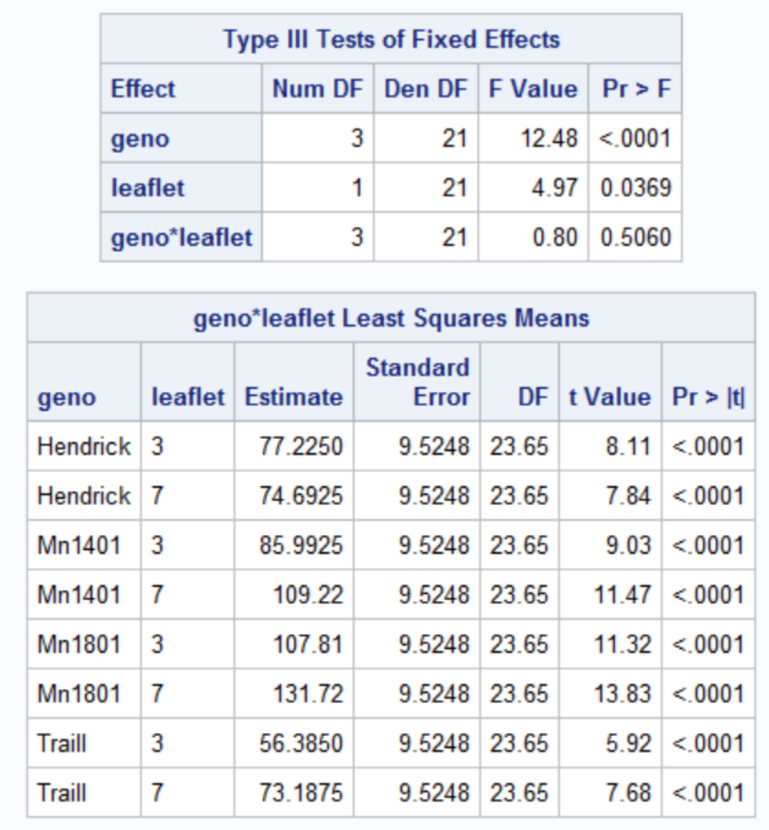
Conclusion:

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.

7. 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.

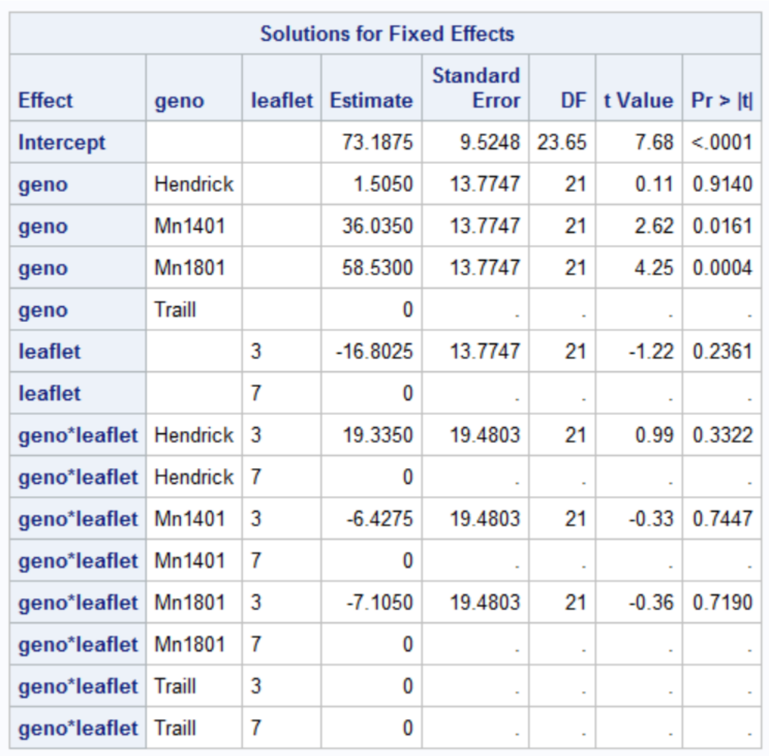
```
TITLE 'marginal model: ';
 PROC GLIMMIX DATA = soy NOBOUND PLOTS = STUDENTPANEL(blup)
outdesign=matrix;
   CLASS block geno leaflet;
   MODEL Area = geno leaflet geno*leaflet /solution ddfm=kr;
   RANDOM _residual_ / solution TYPE = cs SUBJECT = block V;
   LSMEANS geno*leaflet /  PLOT = meanplot(sliceby=leaflet
join cl);
   LSMEANS geno / adjust = tukey;
RUN;
```



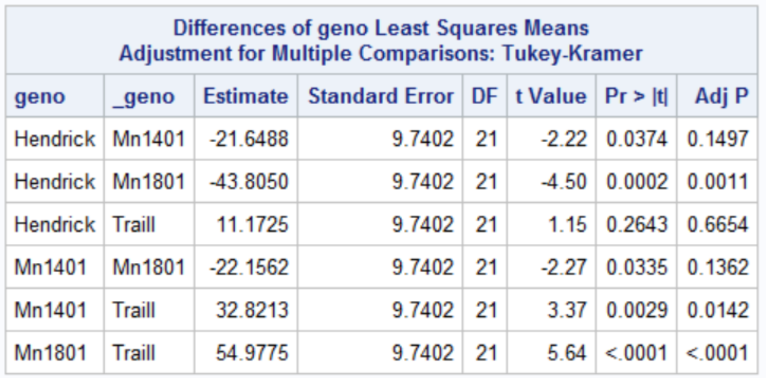


Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
geno	3	21	12.48	<.0001
leaflet	1	21	4.97	0.0369
geno*leaflet	3	21	0.80	0.5060

geno*leaflet Least Squares Means						
geno	leaflet	Estimate	Standard Error	DF	t Value	Pr > |t|
Hendrick	3	77.2250	9.5248	23.65	8.11	<.0001
Hendrick	7	74.6925	9.5248	23.65	7.84	<.0001
Mn1401	3	85.9925	9.5248	23.65	9.03	<.0001
Mn1401	7	109.22	9.5248	23.65	11.47	<.0001
Mn1801	3	107.81	9.5248	23.65	11.32	<.0001
Mn1801	7	131.72	9.5248	23.65	13.83	<.0001
Traill	3	56.3850	9.5248	23.65	5.92	<.0001
Traill	7	73.1875	9.5248	23.65	7.68	<.0001

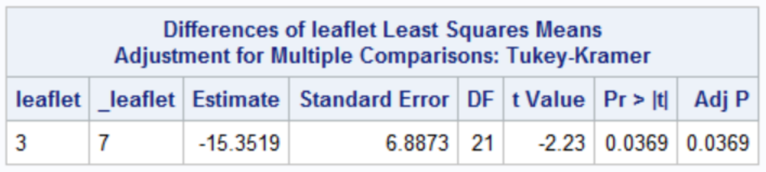


Solutions for Fixed Effects							
Effect	geno	leaflet	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			73.1875	9.5248	23.65	7.68	<.0001
geno	Hendrick		1.5050	13.7747	21	0.11	0.9140
geno	Mn1401		36.0350	13.7747	21	2.62	0.0161
geno	Mn1801		58.5300	13.7747	21	4.25	0.0004
geno	Traill		0	.	.	.	.
leaflet		3	-16.8025	13.7747	21	-1.22	0.2361
leaflet		7	0	.	.	.	.
geno*leaflet	Hendrick	3	19.3350	19.4803	21	0.99	0.3322
geno*leaflet	Hendrick	7	0	.	.	.	.
geno*leaflet	Mn1401	3	-6.4275	19.4803	21	-0.33	0.7447
geno*leaflet	Mn1401	7	0	.	.	.	.
geno*leaflet	Mn1801	3	-7.1050	19.4803	21	-0.36	0.7190
geno*leaflet	Mn1801	7	0	.	.	.	.
geno*leaflet	Traill	3	0	.	.	.	.
geno*leaflet	Traill	7	0	.	.	.	.



Differences of geno Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

geno	_geno	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Hendrick	Mn1401	-21.6488	9.7402	21	-2.22	0.0374	0.1497
Hendrick	Mn1801	-43.8050	9.7402	21	-4.50	0.0002	0.0011
Hendrick	Traill	11.1725	9.7402	21	1.15	0.2643	0.6654
Mn1401	Mn1801	-22.1562	9.7402	21	-2.27	0.0335	0.1362
Mn1401	Traill	32.8213	9.7402	21	3.37	0.0029	0.0142
Mn1801	Traill	54.9775	9.7402	21	5.64	<.0001	<.0001



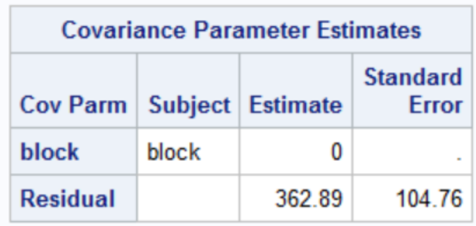
Differences of leaflet Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

leaflet	_leaflet	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
3	7	-15.3519	6.8873	21	-2.23	0.0369	0.0369

8. 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 σ_B^2 to be negative in the conditional model, forcing the covariance matrix of Y to be $\sigma_B^2 I$. The marginal model would not have this problem, and estimates from it remains valid.

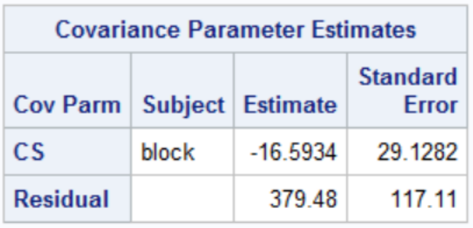
Conditional Model



Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
block	block	0	.
Residual		362.89	104.76

Marginal Model



Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
CS	block	-16.5934	29.1282
Residual		379.48	117.11