HW 4

1. Let’s examine an experiment involving plant species (Species A and Species B) and fertilizer (control, F1 to F3). The plants will be grown in a greenhouse and the outcome of interest is the plant’s height. We want to assign all combinations of fertilizer and species and have 48 pots.
2. What is the experimental unit in this problem? Is there replication? Is it crossed or nested?

Experimental units are plants.

There are 6 replications for each specie-fertilizer combination.

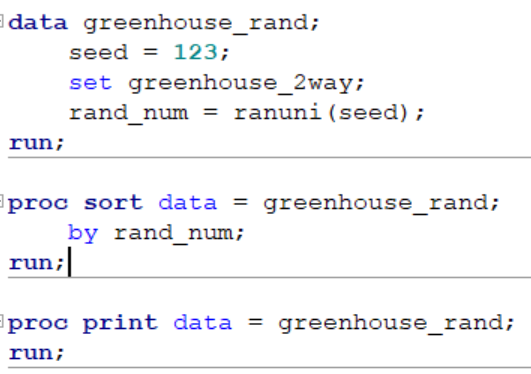
This is a crossed design, because we used all factor-level combinations.

1. Due to the lay out of the greenhouse, we need to arrange the 48 pots as “columns” of 8 pots in 6 “rows”. Design an experiment by assigning the “id” variable (from 4\_homeworkGreenhouseData.sas ) to this drawing. Be very specific how you are doing this assignment.

1. Generate independent uniform random number for each experiment id(use seed for reproducibility.

2. Order experiment settings by the random number generated.

3. Fill the “id” into the pots by row (or column).





1. Analyze the results of this experiment using the data in 4\_homeworkGreenhouseData.sas. Include and interpret a profile plot for both the interaction model and the additive model, as well as a statistical conclusion and scope of inference.

**Assumption Check**

**For interaction model:** Residual vs. predicted value plot shows no obvious pattern, so the constant variance assumption is reasonable. QQ plot of residuals indicate normality assumption is reasonable. There are no high influential values.

**For additive model**: Residual vs predicted value plot reveals increasing variance. Normality assumption seems reasonable. There are no high influential observations.

**Profile plot**

For interaction model, F-test suggests sufficient evidence (p-value = 0.005) for interaction effect. Specie B are taller in average height, and the effect of fertilizer affect the two species differently.

In additive model where no interaction term is specified, the profile plot from SAS would always be parallel. Species B are taller in average height.

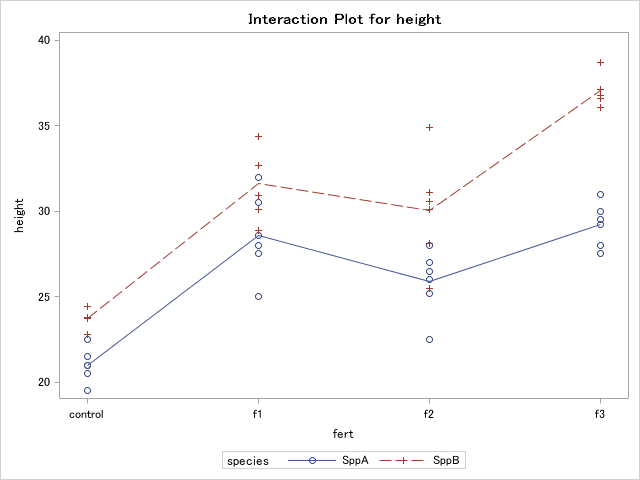
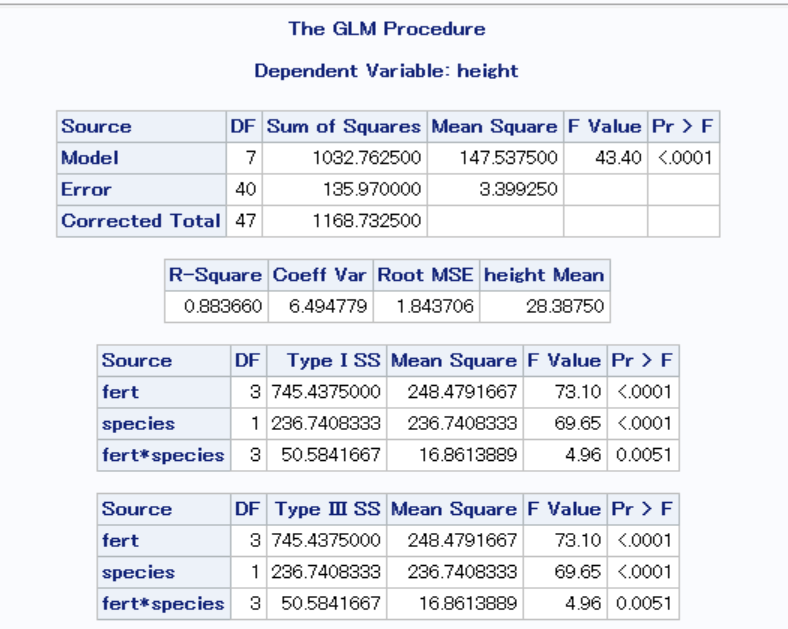
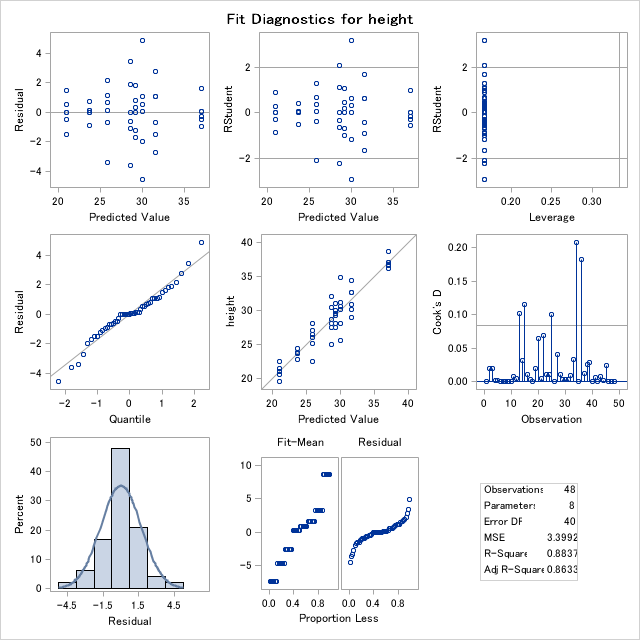
**Conclusion**

Since the interaction term is significant, interaction model is used for inference. There is evidence that treatment with fertilizer caused a significant increase in growth for both plant species compared to controls (two-sided p-values all p<0.05 for treated plant pots vs controls). There was evidence for interaction between fertilizer and plant species (p=0.0051), showing that the fertilizers affect each plant species differently. Plant species B grew significantly taller than plant species A with fertilizers 2 and 3 (two-sided p-values all p<0.05), but there were no significant differences in height between the two species in the control condition and with fertilizer 1 treatment. For plant species A, there were no significant differences in effect on growth between fertilizers. For plant species B, fertilizer 3 caused the greatest growth (p=0.0002 and p<.0001 for differences in growth with fertilizers 1 and 2, respectively). Plant B grew 13.4cm taller than the control with fertilizer 3 with 95% confidence interval [9.8,16.9] cm.

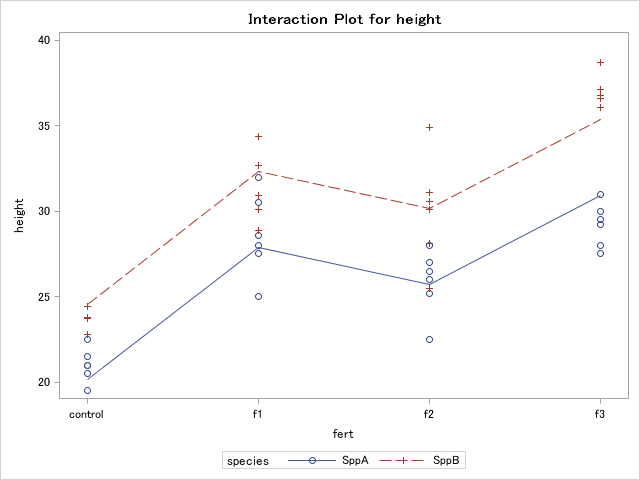
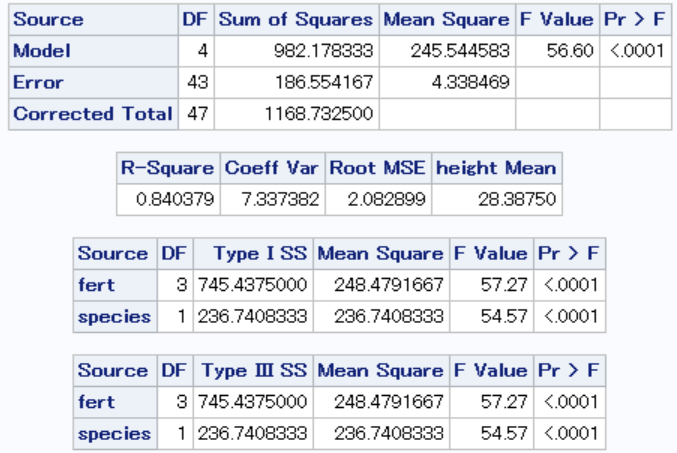
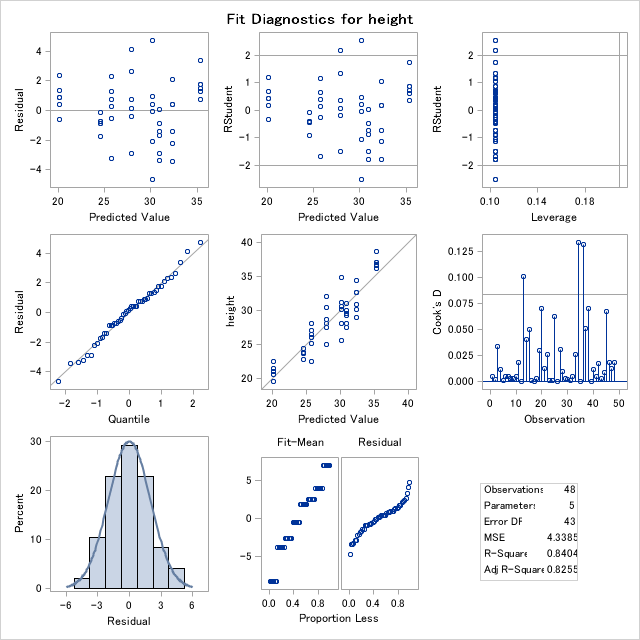
**Scope of inference**

Since this was a controlled randomized experiment, we can conclude that the fertilizers enhanced the plants’ growth. We don’t have information about how plant seeds were chosen, but if they were chosen randomly, we can generalize the effects of the fertilizers to other plants of species A and B.

**Interaction Model**



**Additive Model**



1. Suppose we have the same greenhouse as the previous problem, but that this greenhouse experiment is being done the winter and the surrounding buildings/tree mean that the southern side of the greenhouse gets noticeably more sunshine than the northern side.

a. What is the experimental unit in this problem? Is there replication? Is it crossed or nested?

Experiment units are still plants.

Splitting the pots into southern (row 3-6) and northern (row 1-3) side, there could still be replicates replication. Now each factor-level combination has 3 replications.

This is still a crossed experiment, because each level of one factor appears in all combinations of other factor levels.

b. Design an experiment by assigning the “id” variable (from 4\_homeworkGreenhouseDesign.sas ) to this drawing. Be very specific how you are doing this assignment.

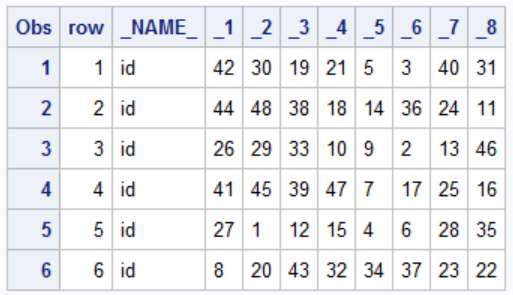
South

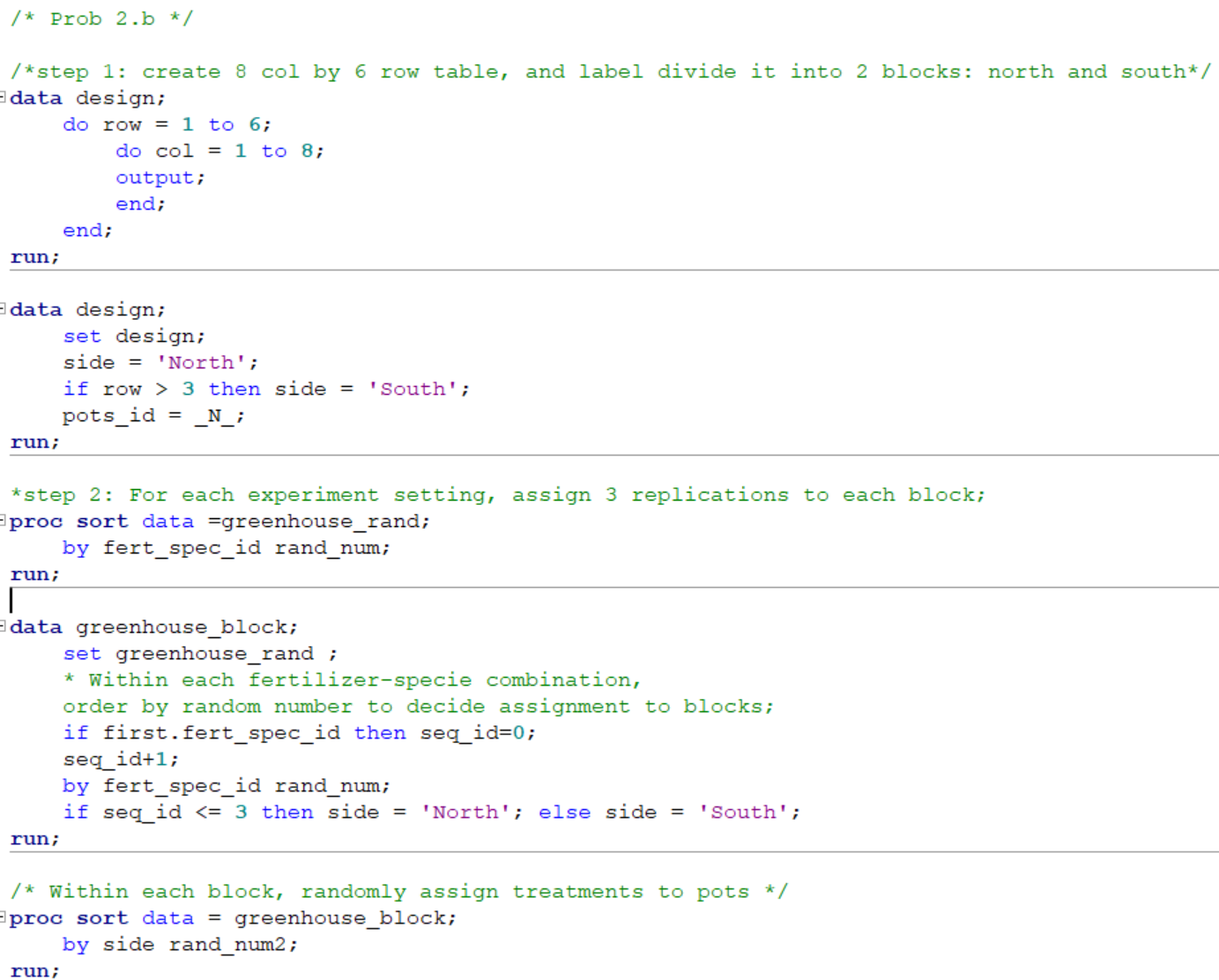
There might be confounding effect due to row locations. To cope with this, we can divide the 48 pots into two blocks: north and south, and randomize fertilizer-species combinations within each block.

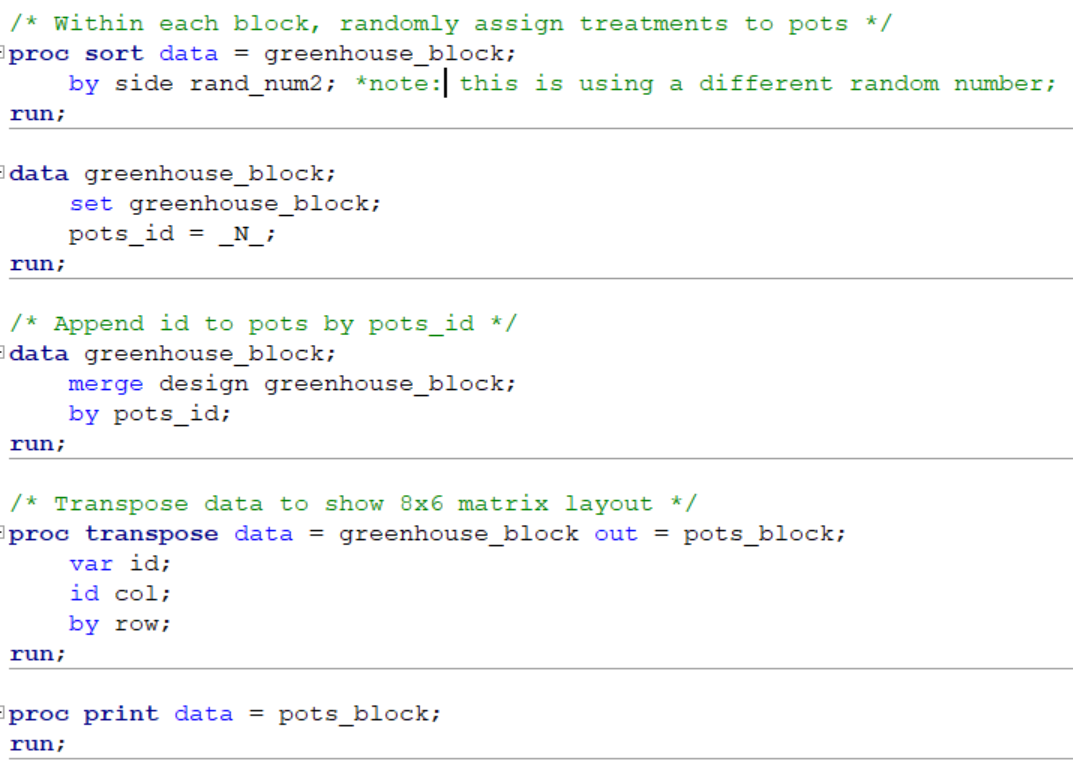
1. Divide the 48 plots into north and south side.

2. Within each of the 8 fertilizer-species settings, randomly assign 3 replications to each block.

See SAS code for how this is achieved.





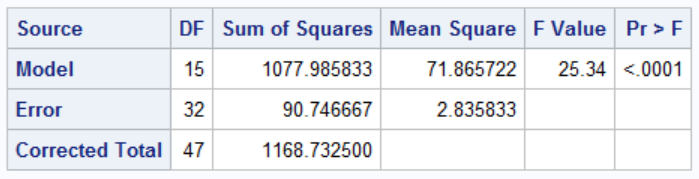


1. Analyze the results of this experiment using the data in 4\_homeworkGreenhouseData.sas, making sure to make any changes to the SAS data set to incorporate any updates to your experiment. Include in this a profile plot for both the interaction model and the additive model, as well as a statistical conclusion and scope of inference.

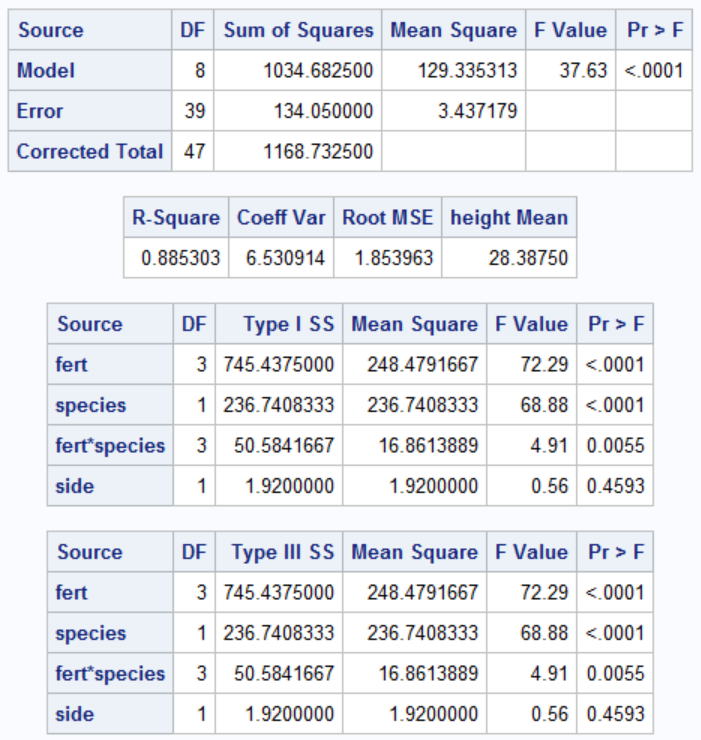
Comparing the full interaction model with fertilizer-species interaction and side block model, F-test (p-value = 0.0628, F-stat = 2.1815, df1 = 7, df2 = 32) indicates the reduced model is sufficient. The fertilizer-species interaction effect is significant after accounting for variability for block factor, therefore conclusions would remain similar to 1.(c) with slightly different estimates.

Note that even for the reduced model, we could choose to keep the factor “side” in the model to account for the variation in data due to it.

ANOVA Table for model: **fert|species|side**

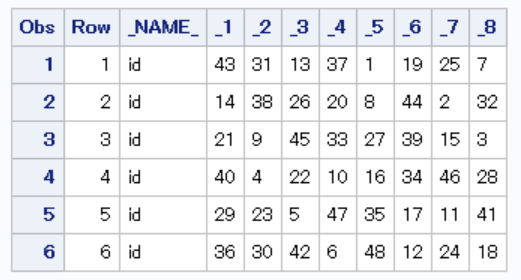


ANOVA Table for model: **fert|species, side**



**Appendix: Additional notes**

Notice that the in Problem 2.(b), we only considered potential confounding in rows and used only one block. If confounding in both rows and columns are of concern, we can use Latin Square Design, which blocks both of those factors. The pots assignment matrix is given below.



/\*HW4: Additional Notes about Latin Square Design \*/

**data** greenhouse\_2way;

input id $ fert $ species $ height;

datalines;

1 control SppA 21.0

2 control SppA 19.5

3 control SppA 22.5

4 control SppA 21.5

5 control SppA 20.5

6 control SppA 21.0

7 control SppB 23.7

8 control SppB 23.8

9 control SppB 23.8

10 control SppB 23.7

11 control SppB 22.8

12 control SppB 24.4

13 f1 SppA 32.0

14 f1 SppA 30.5

15 f1 SppA 25.0

16 f1 SppA 27.5

17 f1 SppA 28.0

18 f1 SppA 28.6

19 f1 SppB 30.1

20 f1 SppB 28.9

21 f1 SppB 30.9

22 f1 SppB 34.4

23 f1 SppB 32.7

24 f1 SppB 32.7

25 f2 SppA 22.5

26 f2 SppA 26.0

27 f2 SppA 28.0

28 f2 SppA 27.0

29 f2 SppA 26.5

30 f2 SppA 25.2

31 f2 SppB 30.6

32 f2 SppB 31.1

33 f2 SppB 28.1

34 f2 SppB 34.9

35 f2 SppB 30.1

36 f2 SppB 25.5

37 f3 SppA 28.0

38 f3 SppA 27.5

39 f3 SppA 31.0

40 f3 SppA 29.5

41 f3 SppA 30.0

42 f3 SppA 29.2

43 f3 SppB 36.1

44 f3 SppB 36.6

45 f3 SppB 38.7

46 f3 SppB 37.1

47 f3 SppB 36.8

48 f3 SppB 37.1

;

**run**;

**data** greenhouse\_2way;

set greenhouse\_2way;

\* Create fert-species combination id;

if fert = 'control' and species = 'SppA' then fert\_spec\_id = **1**;

else if fert = 'f1' and species = 'SppA' then fert\_spec\_id = **2**;

else if fert = 'f2' and species = 'SppA' then fert\_spec\_id = **3**;

else if fert = 'f3' and species = 'SppA' then fert\_spec\_id = **4**;

else if fert = 'control' and species = 'SppB' then fert\_spec\_id = **5**;

else if fert = 'f1' and species = 'SppB' then fert\_spec\_id = **6**;

else if fert = 'f2' and species = 'SppB' then fert\_spec\_id = **7**;

else if fert = 'f3' and species = 'SppB' then fert\_spec\_id = **8**;

**run**;

**proc** **sort** data = greenhouse\_2way;

by fert\_spec\_id;

**run**;

**data** greenhouse\_2way;

set greenhouse\_2way;

/\* creat replication id within each treatment combination,

later used to append to designed matrix;\*/

if first.fert\_spec\_id then seq\_id=**0**;

seq\_id+**1**;

by fert\_spec\_id;

**run**;

/\* Use Latin Square Design that randomize across columns and rows. \*/

title 'Latin Square Design';

**proc** **plan** seed=**123**;

factors Row=**6** ordered Col=**8** ordered / noprint;

treatments Tmt=**8** cyclic;

output out=LatinSquare

Row nvals=(**1** to **6**) random

Col nvals= (**1** to **8**) random

Tmt nvals= (**1** to **8**) random;

**quit**;

**proc** **sort** data=LatinSquare out=LatinSquare;

by Row Col;

**run**;

/\* How treatments are assigned to pots \*/

**proc** **transpose** data = LatinSquare (rename=(Col=\_NAME\_)) out = pots\_tmt;

var tmt;

id col;

by row;

**run**;

**proc** **print** data = pots\_tmt;

**run**;

/\* append original id to designed matrix \*/

**proc** **sql**;

create table assign\_pots as

select a.row

,a.col

,b.\*

from LatinSquare a

left join greenhouse\_2way b

on a.tmt = b.fert\_spec\_id

and a.row = b.seq\_id

order by a.row, a.col;

**quit**;

/\* How original id are assigned to pots \*/

**proc** **transpose** data = assign\_pots out = pots\_id;

var id;

id col;

by row;

**run**;

**proc** **print** data = pots\_id;

**run**;

**data** greenhouse\_side;

set assign\_pots;

if row <= **3** then side = 'north';else side = 'south';

**run**;

**proc** **glm** data = greenhouse\_side plots = all;

class fert species side;

model height = fert|species|side / solution;

**run**;

**proc** **glm** data = greenhouse\_side plots = all;

class fert species side;

model height = fert|species / solution;

**run**;