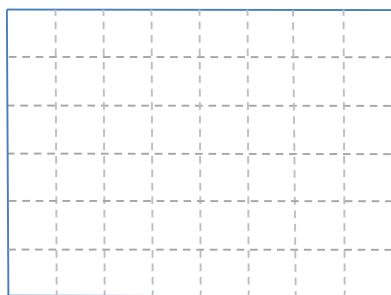


1. Let's examine an experiment involving plant species (Species A and Species B) and fertilizer (control, F1 to F3). The plants will be grown in a greenhouse and the outcome of interest is the plant's height. We want to assign all combinations of fertilizer and species and have 48 pots.
 - a. What is 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.
 - b. 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).

```
data greenhouse_rand;
    seed = 123;
    set greenhouse_2way;
    rand_num = ranuni(seed);
run;

proc sort data = greenhouse_rand;
    by rand_num;
run;

proc print data = greenhouse_rand;
run;
```

	Col_1	Col_2	Col_3	Col_4	Col_5	Col_6	Col_7	Col_8
Row_1	27	47	37	38	9	19	42	3
Row_2	10	6	14	34	32	2	24	5
Row_3	30	8	12	41	25	35	17	33
Row_4	16	36	39	31	28	21	29	40
Row_5	48	23	15	1	43	45	22	11
Row_6	7	20	18	4	26	46	44	13

- c. 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 ($p\text{-value} = 0.005$) for interaction effect. Species 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.

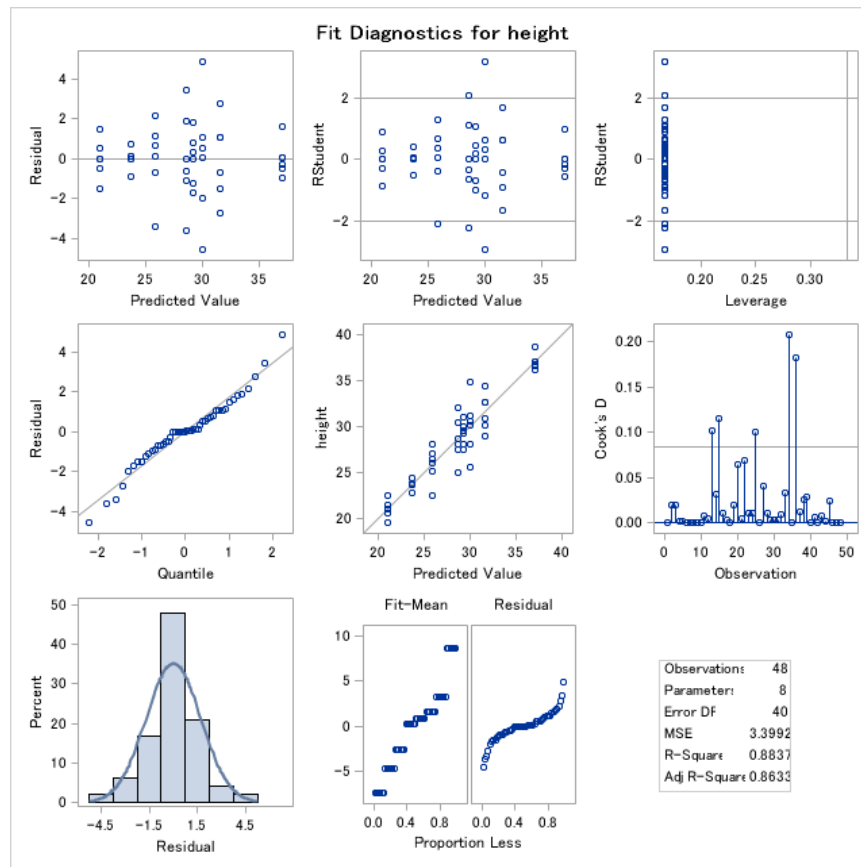
Conclusion

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 $p\text{-values}$ 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 $p\text{-values}$ 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 $p < 0.0001$ 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



The GLM Procedure

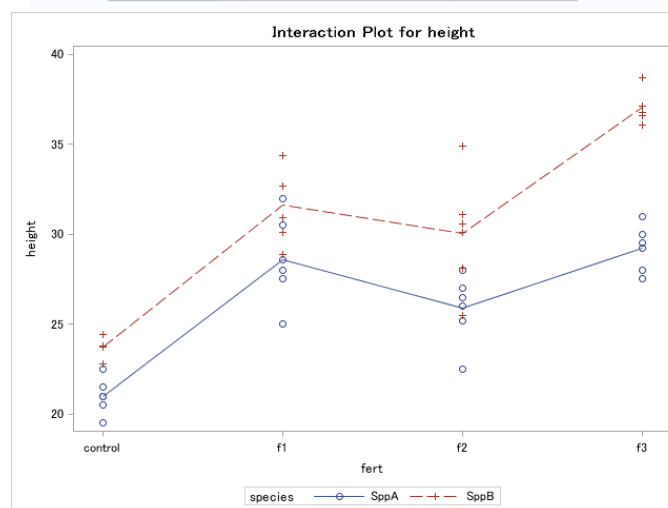
Dependent Variable: height

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1032.762500	147.537500	43.40	<.0001
Error	40	135.970000	3.399250		
Corrected Total	47	1168.732500			

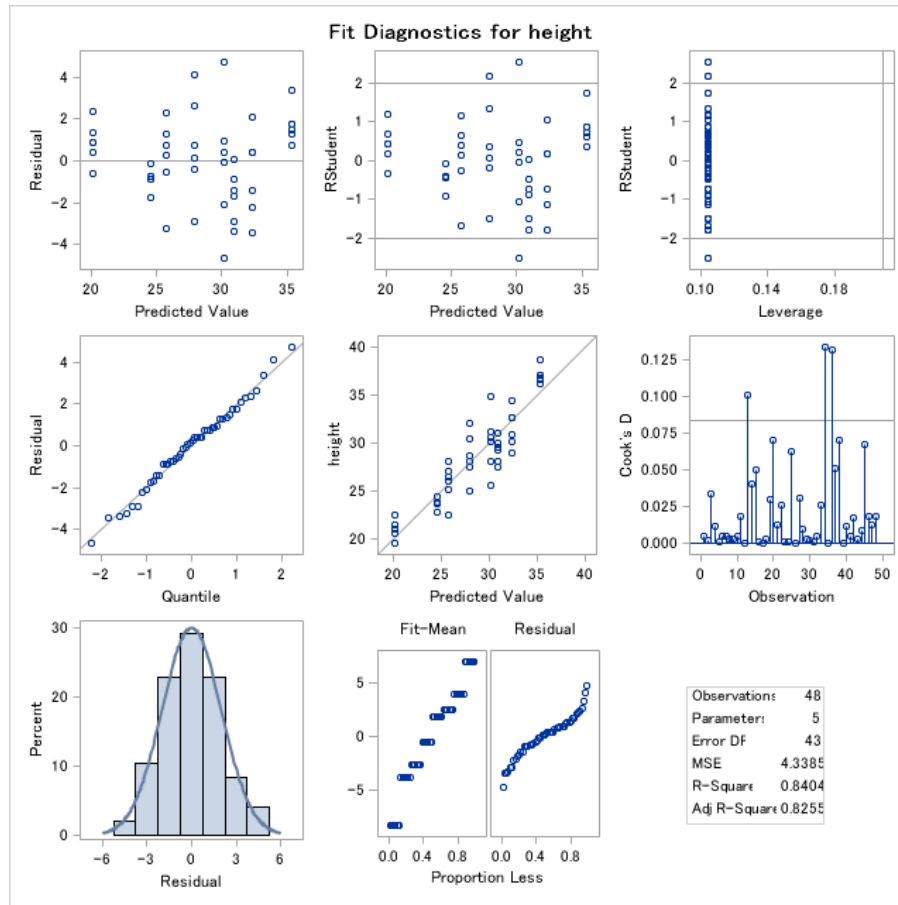
R-Square	Coeff Var	Root MSE	height Mean
0.883660	6.494779	1.843706	28.38750

Source	DF	Type I SS	Mean Square	F Value	Pr > F
fert	3	745.4375000	248.4791667	73.10	<.0001
species	1	236.7408333	236.7408333	69.65	<.0001
fert*species	3	50.5841667	16.8613889	4.96	0.0051

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fert	3	745.4375000	248.4791667	73.10	<.0001
species	1	236.7408333	236.7408333	69.65	<.0001
fert*species	3	50.5841667	16.8613889	4.96	0.0051



Additive Model

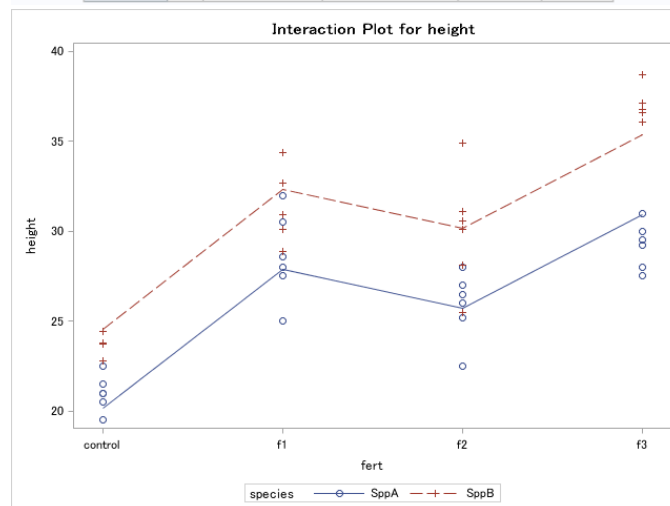


Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	982.178333	245.544583	56.60	<.0001
Error	43	186.554167	4.338469		
Corrected Total	47	1168.732500			

R-Square	Coeff Var	Root MSE	height Mean
0.840379	7.337382	2.082899	28.38750

Source	DF	Type I SS	Mean Square	F Value	Pr > F
fert	3	745.4375000	248.4791667	57.27	<.0001
species	1	236.7408333	236.7408333	54.57	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fert	3	745.4375000	248.4791667	57.27	<.0001
species	1	236.7408333	236.7408333	54.57	<.0001



2. 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.

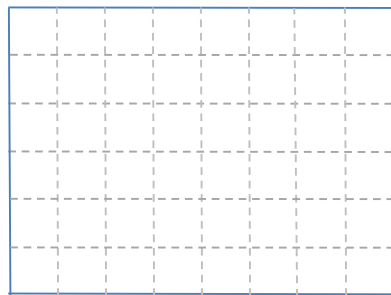
a. What is the experimental unit in this problem? Is there replication? Is it crossed or nested?

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.

Obs	row	_NAME_	_1	_2	_3	_4	_5	_6	_7	_8
1	1	id	42	30	19	21	5	3	40	31
2	2	id	44	48	38	18	14	36	24	11
3	3	id	26	29	33	10	9	2	13	46
4	4	id	41	45	39	47	7	17	25	16
5	5	id	27	1	12	15	4	6	28	35
6	6	id	8	20	43	32	34	37	23	22

```

/* Prob 2.b */

/*step 1: create 8 col by 6 row table, and label divide it into 2 blocks: north and south'
data design;
  do row = 1 to 6;
    do col = 1 to 8;
      output;
    end;
  end;
run;

data design;
  set design;
  side = 'North';
  if row > 3 then side = 'South';
  pots_id = _N_;
run;

*step 2: For each experiment setting, assign 3 replications to each block;
proc sort data =greenhouse_rand;
  by fert_spec_id rand_num;
run;

data greenhouse_block;
  set greenhouse_rand ;
  * Within each fertilizer-specie combination,
  order by random number to decide assignment to blocks;
  if first.fert_spec_id then seq_id=0;
  seq_id+1;
  by fert_spec_id rand_num;
  if seq_id <= 3 then side = 'North'; else side = 'South';
run;

/* Within each block, randomly assign treatments to pots */
proc sort data = greenhouse_block;
  by side rand_num2; *note: this is using a different random number;
run;

data greenhouse_block;
  set greenhouse_block;
  pots_id = _N_;
run;

/* Append id to pots by pots_id */
data greenhouse_block;
  merge design greenhouse_block;
  by pots_id;
run;

/* Transpose data to show 8x6 matrix layout */
proc transpose data = greenhouse_block out = pots_block;
  var id;
  id col;
  by row;
run;

proc print data = pots_block;
run;

```

- c. 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**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	1077.985833	71.865722	25.34	<.0001
Error	32	90.746667	2.835833		
Corrected Total	47	1168.732500			

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

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	1034.682500	129.335313	37.63	<.0001
Error	39	134.050000	3.437179		
Corrected Total	47	1168.732500			

R-Square	Coeff Var	Root MSE	height Mean
0.885303	6.530914	1.853963	28.38750

Source	DF	Type I SS	Mean Square	F Value	Pr > F
fert	3	745.4375000	248.4791667	72.29	<.0001
species	1	236.7408333	236.7408333	68.88	<.0001
fert*species	3	50.5841667	16.8613889	4.91	0.0055
side	1	1.9200000	1.9200000	0.56	0.4593

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fert	3	745.4375000	248.4791667	72.29	<.0001
species	1	236.7408333	236.7408333	68.88	<.0001
fert*species	3	50.5841667	16.8613889	4.91	0.0055
side	1	1.9200000	1.9200000	0.56	0.4593

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.

Obs	Row	_NAME_	_1	_2	_3	_4	_5	_6	_7	_8
1	1	id	43	31	13	37	1	19	25	7
2	2	id	14	38	26	20	8	44	2	32
3	3	id	21	9	45	33	27	39	15	3
4	4	id	40	4	22	10	16	34	46	28
5	5	id	29	23	5	47	35	17	11	41
6	6	id	36	30	42	6	48	12	24	18

```
/*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;
```