IEE 572

Design and Analysis of Experiment

Final Project Report

Growing Tomatoes in Home Garden

By
Shubhra Gupta
Shylaja Kokoori

Instructor
Professor D. C. Montgomery

Problem Statement:

Tomato is one of the most popular vegetables, which is grown at home; however, novice gardeners face lot of problems growing tomatoes, especially from seeds. One of the major issues is that the plants sometimes tend to die if conditions are not favorable. The aim of this project is to find some of the major factors that affect the growth of tomato plants from seeds.

Growing cultivars of tomatoes in containers, with plants still producing the same amount as garden planted varieties, is easy. Therefore, for this project the seeds will be grown in small containers. The biggest advantage to container growing is that the plants can be grown anywhere provided they get at least 6 hours of sunlight. The disadvantage to container growing is that the watering needs to be watched more closely.

Factors, Levels and Range:

There are six main factors, which affect the growth of tomato plant.

- Fertilizer
- Soil
- Seeds
- Water
- Weather
- Area type (Indoor and outdoor)

In this experiment, effect of fertilizer, soil and area type will be studied. These factors were chosen so that the effect of fertilizers and soil types on a given type of seed could be identified. In addition, one batch of seeds will be grown indoor and another outdoors to

see if the area type has any effect. The indoor plants will be kept near a sunny south-facing window, where it can get sunlight abundantly. Fertilizer will be applied once the seeds germinate.

Out of the other factors, weather is a nuisance factor, which is uncontrollable. Seed type will be considering as a constant factor, a variety will be chosen which is suitable for growing in containers. As far as watering is concerned, tomato plants need plenty of water; therefore, the plants will be checked every day in the morning and evening and will be watered, if needed. Plants will be watered slowly near the root and watering will be continued until water starts coming out from the bottom of the container.

This is a 3-factor factorial experiment and two levels of each factor will be considering. The factors, levels and ranges are given below in the table:

Factors	Levels	Range
Fertilizer	- No Fertilizer - Miracle Gro	0 1
Soil	- Top soil and compost - Loam soil (Potting Soil)	
Area type	- Indoor - Outdoor	5-6 hours of sunlight

Response variable:

Response variables that were identified are time of the germination of seeds, height of the plant and number of leaves. Time of germination will be measured in days and height of the plant will be measured in cm/inches using a scale.

Proposed Design for the Experiment:

The design considered for this experiment is a 2^3 factorial design with two replicates resulting in 16 runs. The type of the three factors considered for this experiment is categorical, with two levels high and low. The experiment is blocked based on time, with each replicate forming a block.

Many packets of seeds will be required to run the experiment and variation in seeds is a possibility. Therefore, seed is considered a nuisance factor in this experiment, and is controlled by mixing different packet of seeds.

The containers in which the seeds are grown will be numbered 1 to 16 where first eight runs belong to block1 and the next eight to block2. The containers will be arranged as shown in Figure 1 for first block and similarly for the second block.

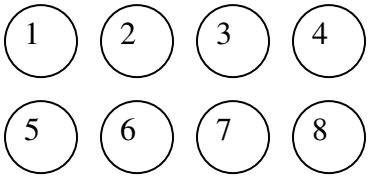


Figure 1

In order to randomize the runs for the experiment, Design Expert Software will be used to determine the run order. Test matrix for the experiment is given in Table 1.

Std	Run	Block	Factor 1 A:Fertilizer	Factor 2 B: Soil	Factor 3 C:Area Type	Response 1 Germination Time	Response 2 Leaf Number	Response 3 Plant Height
5	1	Block 1	A1	B2	C1			
1	2	Block 1	A1	B1	C1			
13	3	Block 1	A1	B2	C2			
3	4	Block 1	A2	B1	C1			
9	5	Block 1	A1	B1	C2			
15	6	Block 1	A2	B2	C2			
7	7	Block 1	A2	B2	C1			
11	8	Block 1	A2	B1	C2			
14	9	Block 2	A1	B2	C2			
16	10	Block 2	A2	B2	C2			
12	11	Block 2	A2	B1	C2			
8	12	Block 2	A2	B2	C1			
4	13	Block 2	A2	B1	C1			
2	14	Block 2	A1	B1	C1			
10	15	Block 2	A1	B1	C2			
6	16	Block 2	A1	B2	C1			

Table 1

Here A1, B1, C1 correspond to low levels of the factors and A2, B2, C2 with the the high levels.

Setup of Experiment:

The experiment was performed in the house of one of the team members to avoid variability due to change in experimenter. A cherry tomato seed variety was chosen as the seed. The experiment was designed to have three factors each at two levels. Therefore as first step two different kinds of soil, namely topsoil and loam soil were placed in the containers at same height in each container. Same numbers of seeds were added to each container, small amount of soil was placed on top of the seeds to cover the seeds, and the additional soil placed was spread out evenly. One set of containers was kept indoors near a south-facing window, and the other set of 8 containers was placed outside where it could get ample sunlight. However since it was very cold outside during the night, the

plants outside were covered with polythene cover. To maintain uniformity the indoor plants also were covered the same way. The containers were watered every day in the evenings. They were watered until water came out from the bottom of the container. The plants were fertilized with 4-day duration.

It was observed that most of the plants indoor germinated within 5 days. Unfortunately, due to the frost, none of the seeds in the container that were placed outside germinated. Because of this problem, the factor corresponding to area type was dropped from the experiment, which converted the experiment to 2-factor factorial with two levels and two replicates. The experiment is still blocked over time as in the original experimental design.

Appropriate germination time was noted down by watching the containers regularly. Other two response variables namely length of plant and number of leaves were recorded slightly more than two weeks from the start of the experiment. Within a container, the germination time for each seed was very different which could be due to the variability in the seeds within a packet or because the seeds were pushed deeper into the soil while watering therefore while recording the response variables the seeds that germinated first are the ones being studied.

Statistical Analysis of the Data:

The experiment was performed according to the Test Matrix given in Table 1. However since the plants placed outside didn't grow, factor corresponding to Area Type was removed and analysis was done based on the data obtained from the other two factors. Test Matrix used is given in Table.2.

Std	Run	Block	Factor 1 A:Fertilizer	Factor 2 B: Soil	Response 1 Germination Time	Response 2 Leaf Number	Response 3 Plant Height
1	1	Block 1	A1	B1			
2	6	Block 2	A1	B1			
3	4	Block 1	A2	B1			
4	8	Block 2	A2	B1			
5	2	Block 1	A1	B2			
6	5	Block 2	A1	B2			
7	3	Block 1	A2	B2			
8	7	Block 2	A2	B2			

Table.2

Response Variable: Leaf Number

ANOVA table for the response variable Leaf Number

Response: Leaf Number

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	0.13	1	0.13			
Model	55.38	3	18.46	40.27	0.0064	significant
A	*55.12*	*1*	*55.12*	*120.27*	*0.0016*	
B	*0.13*	*1*	*0.13*	*0.27*	*0.6376*	
AB	*0.13*	*1*	*0.13*	*0.27*	*0.6376*	
Residual	1.38	3	0.46			
Cor Total	56.88	7				

Std. Dev.	0.68	R-Squared	0.9758
Mean	6.13	Adj R-Squared	0.9515
C.V.	11.05	Pred R-Squared	0.8277
PRESS	9.78	Adeq Precision	10.743

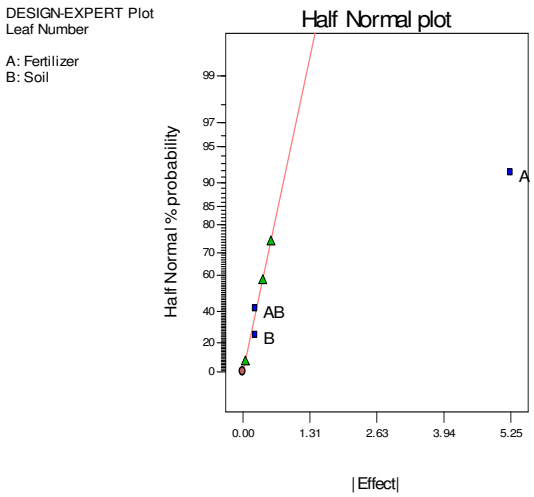
Final Equation in Terms of Coded Factors:

Leaf Number = +6.13 +2.63 * A -0.13 * B -0.12 * A * B

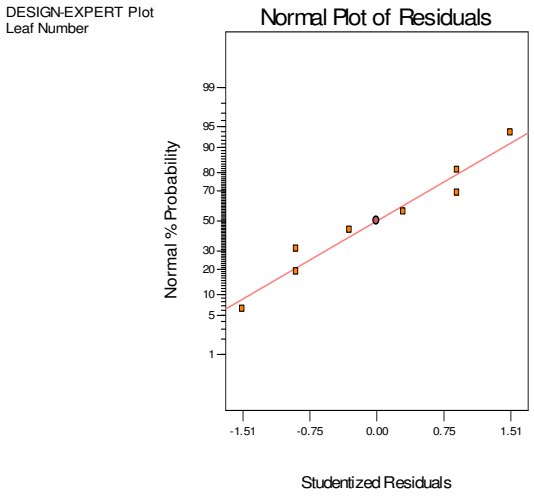
From the ANOVA table it is seen that the full model for the response variable Leaf Number, is significant. The model F-value is large, which implies atleast one variable

has a nonzero effect. Looking at each factorial effect it can be seen that Factor A, fertilizer, is significant and Factor B, Soil, is not significant. From the R-Squared value it can be said that the model account for more than 97% of the variability in the model, which is good. Also the Adj R-Squared value is 95%.

Graph 1: Half Normal Probability Plot for Leaf Number (Full Model)



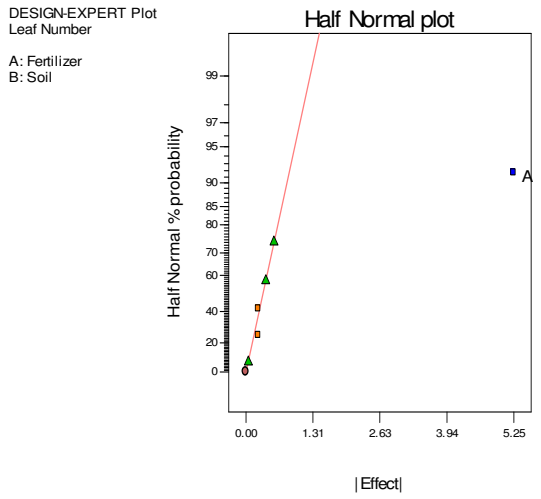
Graph 2: Residual Plot for Leaf Number (Full Model)



From the half normal plot it can be seen that Factor A has a very big effect while factor B and interaction AB does not seem to have a lot of effect. Therefore factor B and the interaction term can be dropped from the model. Residual plot for the data appears as shown in Graph 2. From the residual plot it can be said that the data does not display any abnormality. There is no visible violation of normality.

Inorder to check if dropping factor B and interaction are beneficial, ANOVA for the refined model is considered, where factor B and interaction term AB have been removed. Given below are the half normal plot and the ANOVA table for the refined model.

Graph 3:Half Normal Plot for Leaf Number (Refined Model)



Response: Leaf Number

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

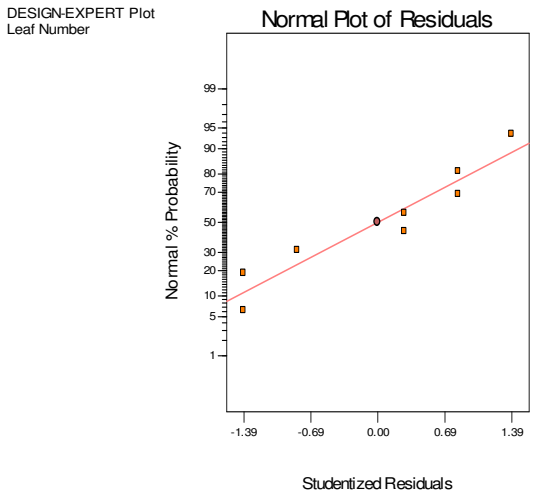
Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	0.13	1	0.13			
Model	55.13	1	55.13	169.62	< 0.0001	significant
A	*55.12*	*1*	*55.12*	*169.62*	*< 0.0001*	
Residual	1.63	5	0.33			
Cor Total	56.88	7				

Std. Dev.	0.57	R-Squared	0.9714
Mean	6.13	Adj R-Squared	0.9656
C.V.	9.31	Pred R-Squared	0.9267
PRESS	4.16	Adeq Precision	15.755

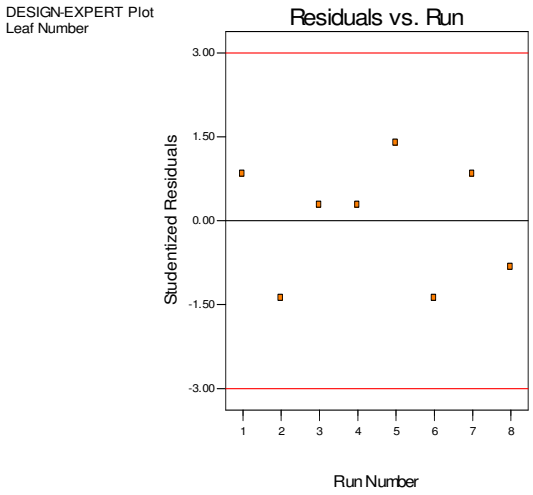
Final Equation in Terms of Coded Factors:
Leaf Number = +6.13 +2.63 * A

From the ANOVA table it is seen that in the refined model for the response variable Leaf Number is significant after dropping the factor and interaction term. The model F-value of 169.62 is high and there is only a 0.01% chance that a model F-value this large could occure due to noise. Adj R-Squared value is improved by more than 1%, which implies that dropping, the factor and interaction term was good. It can also be seen that the PRESS value decreased considerably in the refined model which is desirable. This implies dropping factor B and interaction are favorable. Various plots are considered to see if any of the assumption are violated.

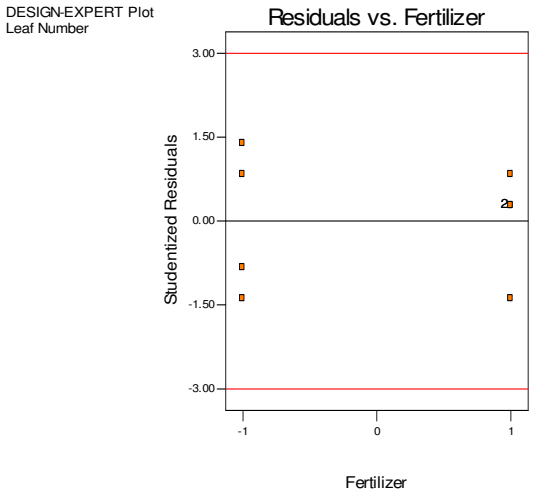
Graph 4:Residual Plot for Leaf Number (Refined Model)

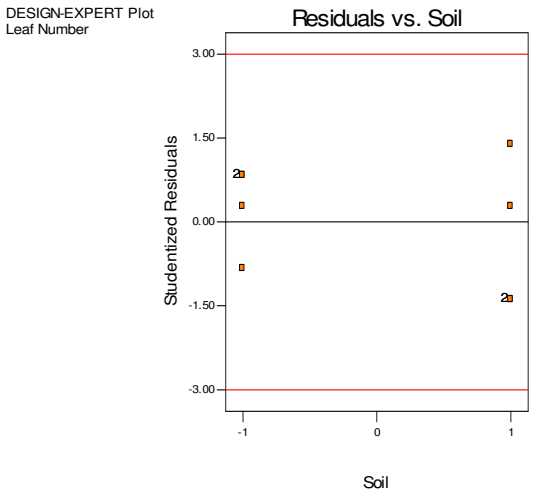


Graph 5: Residual vs Run Order for Leaf Number (Refined Model)

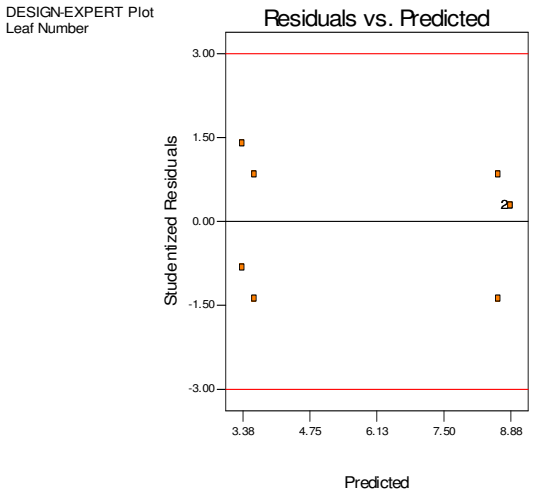


Graph 6: Residual vs Factor for Leaf Number (Refined Model)



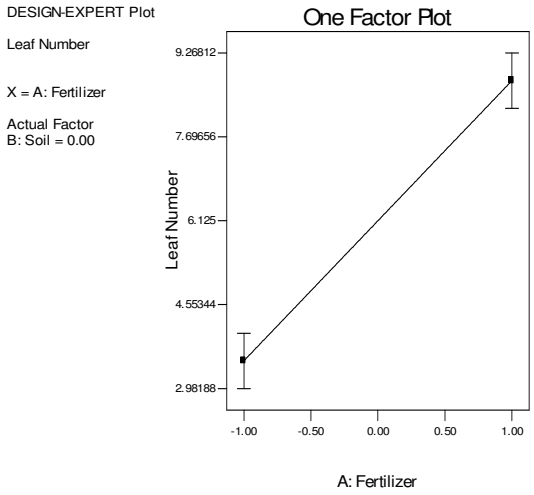


Graph 7: Residual vs Predicted for Leaf Number (Refined Model)



Residual Plot does not display any abnormality also no outliers can be detected. From the Residual Vs Run plot it can be said that there does not seem to be any correlation between the runs hence independence assumption has not been violated.From the Residual Vs Factor graphs it can be said that there is some change in the variance in both the graphs. Also Residual Vs Predicted graph again shows some anomaly in constant variance assumption. However it doesn't look very bad. Next the model graphs are studied.

Graph 8: One Factor Plot for fertilizer



This plot shows that the leaf number increased with application of fertilizer.

Thus analysis of data obtained for the response variable leaf number it can be said that no assumption are violated and application of fertilizer is favorable in this case, though soil does not play a very important role here.

Response Variable: Plant Height

ANOVA table for response variable plant height.

Response: Plant Height

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	0.045	1	0.045			
Model	3.81	3	1.27	3.72	0.1547	not significant
A	*1.81*	*1*	*1.81*	*5.28*	*0.1051*	
B	*2.00*	*1*	*2.00*	*5.85*	*0.0942*	
AB	*5.000E-003*	*1*	*5.000E-003*	*0.015*	*0.9114*	
Residual	1.03	3	0.34			
Cor Total	4.88	7				

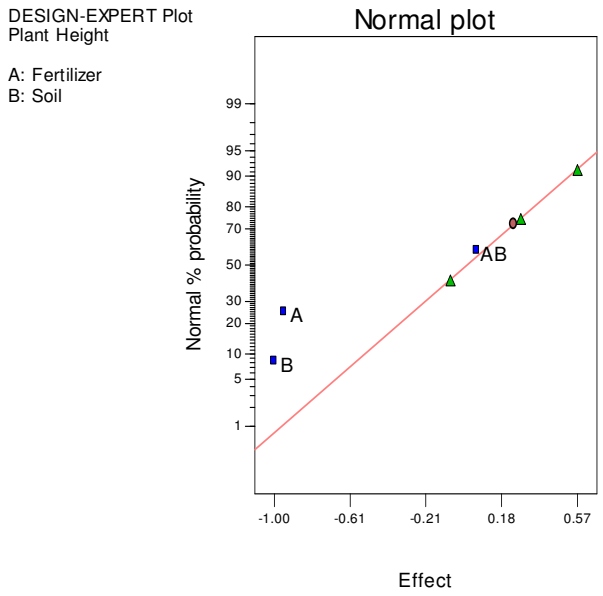
Std. Dev.	0.58	R-Squared	0.7880
Mean	3.65	Adj R-Squared	0.5760
C.V.	16.01	Pred R-Squared	-0.5075
PRESS	7.29	Adeq Precision	4.544

Final Equation in Terms of Coded Factors:

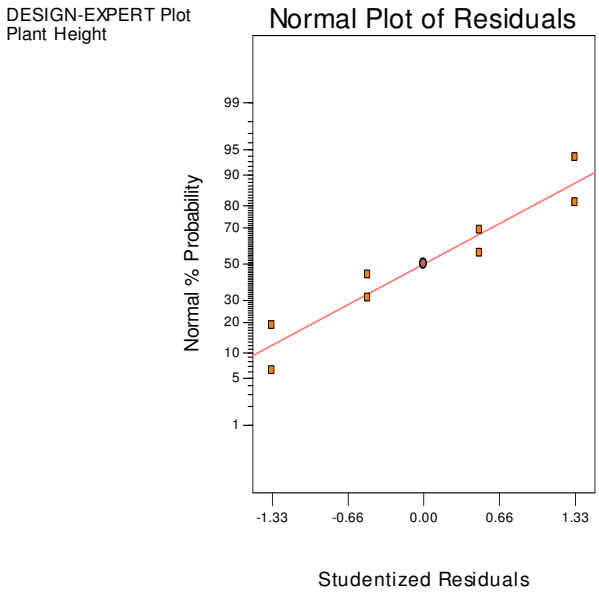
Plant Height = +3.65 -0.48 * A -0.50 * B +0.025 * A * B

From the ANOVA table it is seen that for the response variable Plant Height the full model is not significant. However Factors A & B are significant at 10% & 9% level and it has a big interaction term. From the R-Squared value it can be said that the model account for approx 79% of the variability in the model. Also the Adj R-Squared value is 57.6%, which is small.

Graph 9: Normal Probability Plot (Full Model)



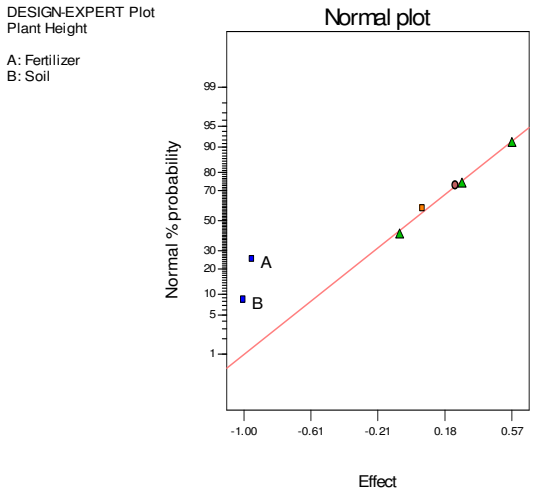
Graph 10: Residual Plot (Full Model)



The normal probability plot for Plant Height (Graph 9) shows large effects that emerge from the main factors. Interaction term is lying along line, and its effect could be is negligible. This supports conclusion made based on ANOVA table. Residual plot for the full model is given in Graph 10. Residual plots do not display any severe indication of abnormality, nor do they display any evidence pointing to possible outliers.

The p-value for the interaction term in the ANOVA table is very high therefore as a next step the model is refined by dropping the interaction term.

Graph 11:Normal Probability Plot for Plant Height (Refined Model)



The ANOVA table for refined model is given below.

Response: Plant Height

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	0.045	1	0.045			
Model	3.80	2	1.90	7.39	0.0454	significant
A	*1.81*	*1*	*1.81*	*7.01*	*0.0571*	
B	*2.00*	*1*	*2.00*	*7.77*	*0.0495*	
Residual	1.03	4	0.26			
Cor Total	4.88	7				

Std. Dev.	0.51	R-Squared	0.7870
Mean	3.65	Adj R-Squared	0.6805
C.V.	13.90	Pred R-Squared	0.1479
PRESS	4.12	Adeq Precision	5.853

Final Equation in Terms of Coded Factors:

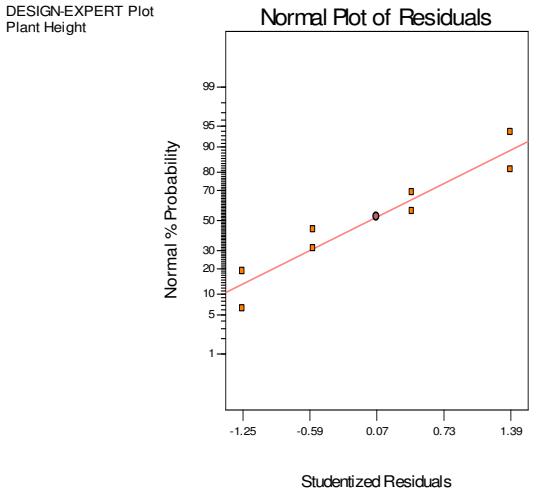
Plant Height = +3.65 -0.48 * A -0.50 * B

The refined model after dropping the interaction term is significant. Factor A is significant at 5.7% level, based on this it cannot be said that fertilizer is not an important factor. Recommendation would be to obtain more data before making a decision

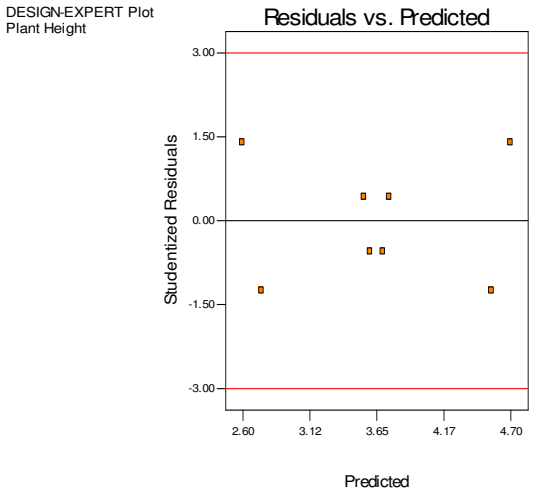
regarding this. Factor B is significant at 4.9%. The model F-value of 7.39 implies the model is significant. There is only a 4.54% chance that a model F-Value this large could occur due to noise.

In the refined model there is not much difference in the R-squared value however the Adj R-Squared has increased which implies dropping the interaction term was good. It can also be seen that the PRESS value decreased in the refined model. All this suggest that dropping the interaction term is favorable. Different plots are considered to see if any of the assumption are violated.

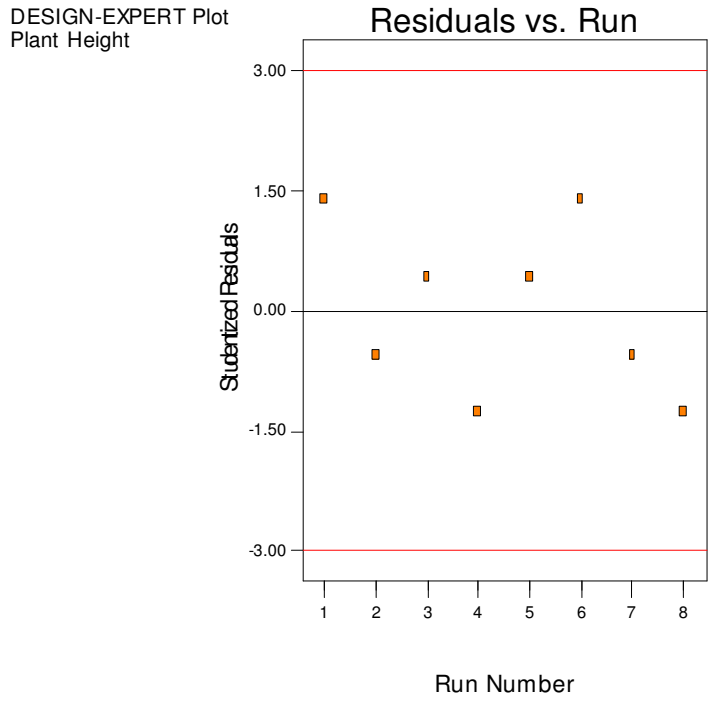
Graph 12: Residual plot for Plant height (Refined Model)



Graph 13:Residual Vs Predicted for Plant height (Refined Model)



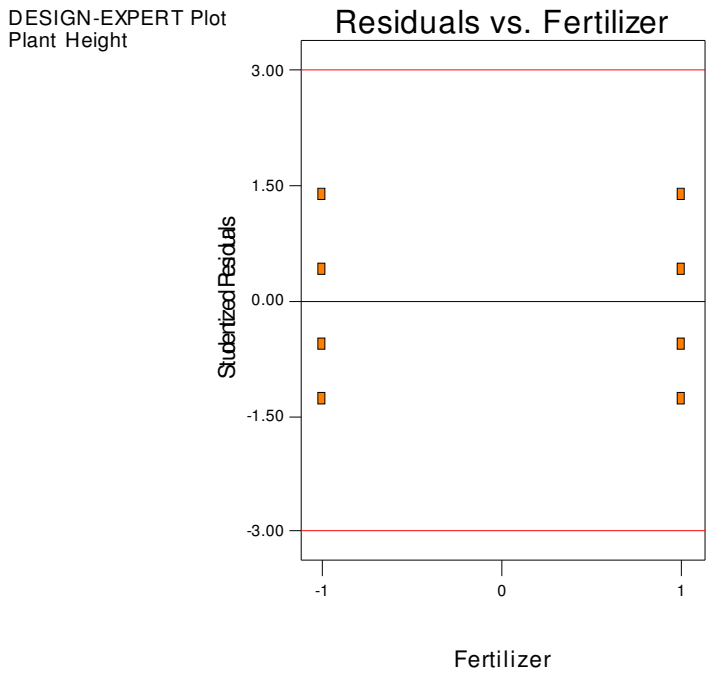
Graph 14:Residual Vs Run for Plant height (Refined Model)



From the residual plot (Graph 12) it can be said that normality assumption has not been violated. It does display some pattern but that is because of low number degrees of freedom for the residuals. This pattern can be observed in the Residual Vs Predicted graph as well. However this can be ignored since the number of runs are small and the

degrees of freedom for the residual is only 3. Also from the Residual vs Run graph(Graph 14) it can be said that independence assumption is not violated.

Graph 15:One factor plot Residual Vs Fertilizer (Refined Model)



Graph 16:One Factor plot Residual Vs Soil for Plant height (Refined Model)

Graph 15 & Graph 16 shows that constant variance assumption not violated.

Response Variable:Germination Time

ANOVA table for response variable Germination Time

Response: Germination Time

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	10.13	1	10.13			
Model	6.38	3	2.13	1.00	0.5000	not significant
A	*0.13*	*1*	*0.13*	*0.059*	*0.8240*	
B	*6.12*	*1*	*6.12*	*2.88*	*0.1881*	
AB	*0.13*	*1*	*0.13*	*0.059*	*0.8240*	
Residual	6.37	3	2.12			
Cor Total	22.88	7				

Std. Dev.	1.46	R-Squared	0.5000
Mean	6.13	Adj R-Squared	0.0000
C.V.	23.80	Pred RSquare	-2.5556
PRESS	45.33	Adeq Precision	3.688

Final Equation in Terms of Coded Factors:

Germination Time = +6.13 +0.12* A +0.87 * B -0.12* A * B

From the ANOVA table it is seen that the full model for the response variable Germination Time is not significant and also dropping any factor is not improving the model. The reason could be that fertilizer was not added before seed germination and also range for soil was not good enough. The model F-value of 1.00 implies the model is not significant relative to the noise. There is a 50.00 % chance that a model F-value this large could occur due to noise. From the R-Squared value it can be said that the model account only for 50% of the variability and also the Adj R-Squared value 0%. Value of PRESS is very high 45.33. Therefore this is not good model.

Therefore the analysis made for this response variable is not being considered in this experiment.

Conclusions and recommendations:

The model is significant for response variables leaf number and plant height. From the analysis it could be said that fertilizer plays an important role in the development of the plant as can be seen by the increase in number of leaves when fertilizer is applied. As far as plant height is concerned soil plays an important role.Effect of fertilizer cannot be determined from the data available.

One of the problems associated with this project could be that soil was not varied over a wider range.Also not thinking through the entire process,especially growing plants in winter, this being an agriculture experiment, is a major weakness in the performance of this experiment.

We would like to repeat the experiment in spring or summer to study the effect of area type.If this experiment is done again, our suggestion would be to grow the plants for a longer duration to learn the extent to which fertilizer affects the growth of the plant. Also we would suggest varying the soil factor over a wider range and more number of levels. There are all kinds of soil available in the market which might not be very different from one another in quality but may differ to a larger extent in price. This method could help to identify a quality product with lower price.