Factorial Analysis of Factors Impacting Tomato Germination and Growth

Macie Hanneken

**Introduction**

There are a variety of factors that can effect growth of fruit and time to germination in any plant, such as weather, temperature, nutrient level, and water to name a few. In this experiment, I am interested in determining if days between watering, the addition of tomato feed to the water used, or soil type has an effect on the seedlings time to germination or the height of the seedlings at the end of the experiment. For this particular experiment, I used tomato seeds due to their quick predicted germination time. For the analysis, I looked at the effect of watering the seeds and seedlings 0 days apart, 1 day apart, or 2 days apart. Along with watering days, I was also curious as to if the addition of tomato feed would have an effect on the time to germination or the seedling height at the end of the experiment. Lastly, I looked at the effect that soil type, organic or non-organic had on the seeds time to germination and the seedling height at the end of the experiment.

**Methods**

There was a total of 72 different seeds used in this experiment. Each seed was planted in a 2x2in square in a planter that had 36 total squares. Two planters were used. To ensure randomization, I randomly assigned each treatment a number and each plot in the planters a number. Then I matched the numbers and put the treatment number corresponding to the plot number in that particular plot. I used Pro-Mix organic potting mix and Schultz Premium Potting Soil Plus for the non-organic soil. I first filled each square in the planter with one-third cup soil. I then placed one seed per each square in the planter and covered each square with the designated soil. I then watered each square with three table spoons of water to ensure that the soil was not too dry and would not inhibit seedling germination. For the treatments that received tomato feed, I used Burpee Organic, water soluble tomato and vegetable feed. I put a half of a teaspoon in a gallon of water and then watered the squares that needed to receive the tomato feed with that once a week after the initial watering. All plots were initially watered with three tablespoons of water and then for the duration of the experiment were watered with two tablespoons of water on their respective watering days. Both planters were placed in front of my balcony door where the blinds remained open in order to ensure the most sunlight. Planters where checked once a day around 6pm for new germination. The experiment lasted four weeks, and the height of the seedlings that sprouted was taken and recorded at the end of the experiment. Each of the planters was turned to face the other direction once a week in order for each square to get an equal amount of sunlight.

*Do different growth factors, days to watering, tomato feed, soil type, or their interactions have an effect on days to germination?*

To determine the effects that these different growth factors and their interactions have on the number of days to germination of the tomato seeds, I will use a fixed effects factorial analysis of variance (ANOVA). The dependent variable will be days to germination and the independent variables will be the days to watering, absence or presence of tomato feed, and the soil type. For the ANOVA, we will use the following model to determine the effect that these factors are having on the tomato seedlings height: . Where is individual observations for *l* replicates, is the true mean growth, is the fixed effect of the soil type, is the fixed effect of the days between watering, is the fixed effect of the addition of tomato feed, is the interaction between the soil type and the days between watering, is the fixed effect of the interaction between soil type and addition of tomato feed, is the fixed effect of the interaction between days between watering and the addition of tomato feed, is the fixed effect of the interaction between the soil type, days between watering, and the addition of tomato feed, and lastly, is the random variation among individuals with the same treatments The null hypothesis is that there will be no effect of days to watering, absence or presence of tomato feed, the soil type, or their interactions. We will reject the null hypothesis if the F test produces a p-value that is smaller than .05. If the null hypothesis is rejected, we can consider the alternative hypothesis. The alternative hypothesis is that at least one of the levels of these treatments will have a significant effect on the days to germination for the tomato seedlings.

Before we can proceed with the factorial ANOVA, four assumptions must be met. The first assumption is that the sampling of all individuals must be random and independent. This is met from the design of the experiment. Assumption two states that the error variances must be independent, this simply means that the mean and the variance must not be correlated. The best way to assure that they are independent, and not correlated is to plot the predicted values vs. the residuals (Figure 1). It appears that as the predicted values get larger, the residuals do as well. This is not good. We are going to need a transformation of the data in order for this assumption to be met. I chose to take the natural log of the days to germination in order to meet the assumption of independence of error variances. Figure 2 shows the plot of the predicted vs. the residual values for the transformed data. The data points are scattered about with no general pattern, and more importantly, as the predicted value goes up, the residuals do not, this implies that they are not correlated and independent, and assumption two is met. The third assumption is the homogeneity of variances. We will refer to Figure 2 (the residual vs. predicted plot for the transformed data) for this as well. If the variances are homogeneous, the data in the plot will be well spread out and not curved, bell shaped, or lumped in anyway. Looking at Figure 2, we can conclude that the plot evenly spread and consider this assumption met. The fourth and final assumption requires the residuals to be normally distributed. We will use the Shapiro-Wilks test to determine if the data is normally distributed. The results are in Table 1 for the transformed and untransformed data. Since the test statistic is close to one and the p-value is greater than .05, we can conclude that residuals are normally distributed for the natural log transformed data. For a visual reference, see Figure 3b. The probability plot’s data points are all close to the regression line, and therefore, the residuals are normally distributed and the fourth assumption is met. We can now do the fixed-effects factorial ANOVA for this data.

*Do different growth factors, days to watering, tomato feed, soil type, or their interactions have an effect on seedling height?*

To determine the effects that these different growth factors and their interactions have on the height of the tomato seedlings, I will use a fixed effects factorial analysis of variance (ANOVA). The dependent variable will be the height of the seedlings and the independent variables will be the days to watering, absence or presence of tomato feed, and the soil type. For the ANOVA, we will use the following model to determine the effect that these factors are having on the tomato seedlings height: . Where is individual observations for *l* replicates, is the true mean growth, is the fixed effect of the soil type, is the fixed effect of the days between watering, is the fixed effect of the addition of tomato feed, is the interaction between the soil type and the days between watering, is the fixed effect of the interaction between soil type and addition of tomato feed, is the fixed effect of the interaction between days between watering and the addition of tomato feed, is the fixed effect of the interaction between the soil type, days between watering, and the addition of tomato feed, and lastly, is the random variation among individuals with the same treatments The null hypothesis is that there will be no effect of days to watering, absence or presence of tomato feed, the soil type, or their interactions. We will reject the null hypothesis if the F test produces a p-value that is smaller than .05. If the null hypothesis is rejected, we can consider the alternative hypothesis. The alternative hypothesis is that at least one of the levels of these treatments will have a significant effect on the tomato seedlings height.

Before we can proceed with the factorial ANOVA, four assumptions must be met. The first assumption is that the sampling of all individuals must be random and independent. This is met from the design of the experiment. Assumption two states that the error variances must be independent, this simply means that the mean and the variance must not be correlated. The best way to assure that they are independent, and not correlated is to plot the predicted values vs. the residuals (Figure 4). In Figure 4, we can see that there is no correlation between mean and error. The data points are scattered about with no general pattern, and more importantly, as the predicted value goes up, the residuals do not, this implies that they are not correlated and independent, and assumption two is met. The third assumption is the homogeneity of variances. We will refer to Figure 4 for this as well. If the variances are homogeneous, the data in the plot will be well spread out and not curved, bell shaped, or lumped in anyway. Looking at Figure 4, we can conclude that the plot evenly spread and consider this assumption met. The fourth and final assumption requires the residuals to be normally distributed. We will use the Shapiro-Wilks test to determine if the data is normally distributed. The results are in Table 2. Since the test statistic is close to one and the p-value is greater than .1, we can conclude that residuals are normally distributed. For a visual reference, see Figure 5. The probability plot’s data points are all close to the regression line, and therefore, the residuals are normally distributed and the fourth assumption is met. We can now do the fixed-effects factorial ANOVA for this data.

**Results**

*Do different growth factors, days to watering, tomato feed, soil type, or their interactions have an effect on days to germination?*

The results from the fixed effects factorial analysis of variance (ANOVA) can be found in Table 3. Table 3 shows that the only treatment or interaction that had a significant effect on the number of days until germination of the tomato seedling was the interaction between tomato feed and soil with a p-value from the F-test less than .05. This means that all other treatments and interactions had no effect on days until germination. Figure 6 shows that the non-organic soil with the feed had seeds germinate in the shortest period of time with the organic without tomato feed coming in at a close second. The non-organic soil without tomato feed took the longest amount of time to germinate. Figure 7 shows the interaction plot for the soil type and absence and presence of tomato feed. If the lines in this plot were parallel, there would be no interaction. The lines cross though, implying that there is an interaction and the analysis tells us that it is significant.

Since the interaction was significant for feed and soil type, I also conducted fixed effects ANOVA on the days to germination for just these two factors and nothing else, to see if the interaction was making it hard to see what might be going on individually with these two factors. The results of this reduced model can be found in Table 4. Neither the type of soil or the absence or presence of tomato feed was significant in the reduced model just as in the full model.

*Do different growth factors, days to watering, tomato feed, soil type, or their interactions have an effect on seedling height?*

The results from the fixed effects factorial analysis of variance (ANOVA) can be found in Table 4. Table 4 shows that all of the p-values for the F-test are greater than .05 and therefore we fail to reject any of the null hypotheses and conclude that none of the treatments or their interactions are having an effect on tomato seedling height.

**Conclusions**

I found that the only treatment or interaction that had an effect on days to germination of the tomato seeds was the interaction between soil type and tomato feed. This showed that tomato seedlings in non-organic soil with tomato feed had the shortest number of days until germination. The tomatoes in the non-organic soil with no feed had the largest number of days until germination. If you were a gardener and you wanted your plants to sprout as quick as possible, this study shows that it would be useful to use non-organic soil with tomato feed to make that happen. In this study though, there were a lot of missing data values from seeds that did not germinate at all. Only seven out of the thirty-six seeds in the non-organic soil sprouted. Since these seeds did not germinate, they were not used in the experiment, but it leads me to believe that the organic soil is better for the seeds since thirty-five out of the thirty-six seeds did sprout in the organic soil. Reasons that these seeds might not have germinated could be due to a variety of factors, such as the non-organic soil had larger chunks of mulch and other things and when the planters were watered the seeds may have got washed to the bottom of the planter making them incapable of growing. It could have been a bad bag of potting mix or just a lot of bad seeds as well, which seems less likely.

Most of the results from this experiment were insignificant, something I think that could be due to the lack of germination of over half on the non-organic plots. I also chose to have a variety of watering days in order to induce some sort of stress on the plants that were watered more days a part. This was not achieved because I believe watering the individual squares two tablespoons of water was too much water. This experiment was extremely short term as well. I would like to conduct a long term experiment under these settings and look at factors such as fruit size and production and seed production under these factors. It would also be beneficial to conduct a study like this in a greenhouse versus an apartment, but resources were limited.

**Appendix**

Plot of resfinal\*prefinal. Legend: A = 1 obs, B = 2 obs, etc.

resfinal |

|

8 +

|

|

| A

|

|

|

6 +

| A

|

|

|

|

|

4 + B

|

|

|

|

|

|

2 +

|

| A

| F

| A

|

|

0 + G F

|

| B

| C D

|

|

|

-2 + B A

| A

|

| C

|

|

|

-4 +

| A

|

|

|

|

|

-6 +

|

---+-------------+-------------+-------------+-------------+-------------+-------------+-------------+-------------+--

9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5

prefinal

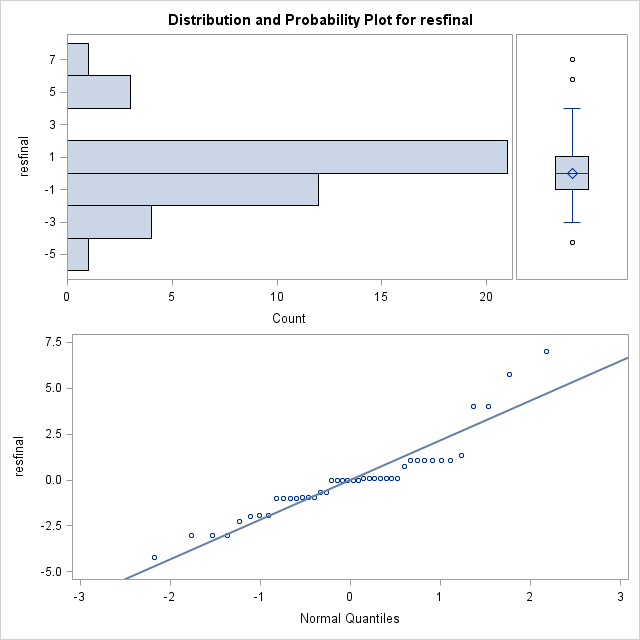
NOTE: 30 obs had missing values.

Figure 1. Plotting predicted values against residual values from the data. This will give a visual representation as to if mean and error are correlated and if the variances are equal

|  |
| --- |
| The SAS System |

|  |
| --- |
| Plot of trresfinal\*trprefinal. Legend: A = 1 obs, B = 2 obs, etc.    trresfinal |  |  0.6 +  |  |  |  | A  0.5 +  |  |  |  |  0.4 + A  |  |  | B  |  0.3 +  |  |  |  |  0.2 +  |  |  | A  |  0.1 + F A  |  |  |  | G F  0.0 +  |  |  | B  | D  -0.1 + C  |  | A  |  | A  -0.2 +  | B  |  |  |  -0.3 + C  |  | A  |  |  -0.4 +  |  ---+---------------+---------------+---------------+---------------+---------------+---------------+--  2.25 2.30 2.35 2.40 2.45 2.50 2.55    trprefinal    NOTE: 30 obs had missing values. |

Figure 2. Plotting predicted values against residual values from the natural log transformed data. This will give a visual representation as to if mean and error are correlated and if the variances are equal.



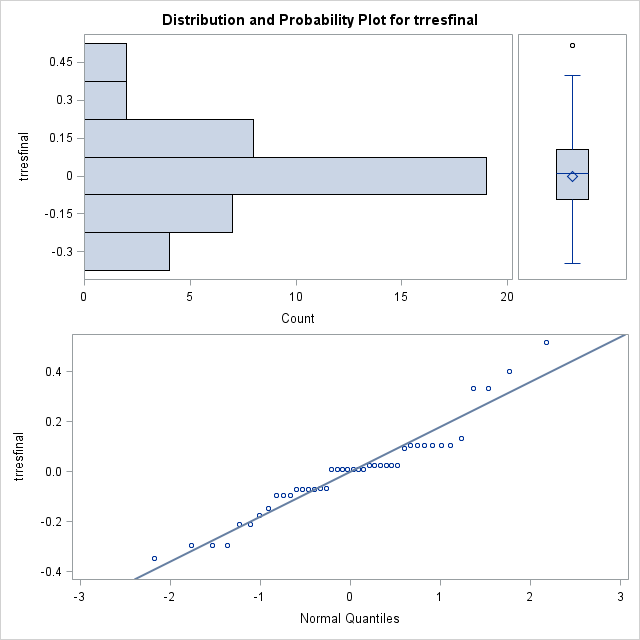


Figure 3. This is the probability plot. 3a (top) shows the normality QQ-plot for the untransformed data. For this plot, the tails are pretty bad and leads us to believe that the residuals are not normally distributed. 3b (bottom) this is the plot for the natural log transformed data. Since this plot is linear and the data points are close to the regression line, we can conclude that the data is normal.

|  |
| --- |
| The SAS System |

|  |
| --- |
| Plot of resfinal\*prefinal. Legend: A = 1 obs, B = 2 obs, etc.    resfinal |  |  5 +  |  |  | A  |  4 +  | A  |  |  |  3 +  | A  | A  | A  | A  2 +  | A  | AA  | A  | A  1 + A A  | A A  | B  | A  |  0 + C A A A A  | A A A  | A B  | A  | A  -1 + A  | A  |  | A  | A  -2 +  | A  |  |  | A  -3 + A A  | A  |  | A  |  -4 +  |  |  | A  |  -5 +  |  ---+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+--  3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0    prefinal    NOTE: 28 obs had missing values. |

Figure 4. Plotting predicted values against residual values from the data. This will give a visual representation as to if mean and error are correlated and if the variances are equal.

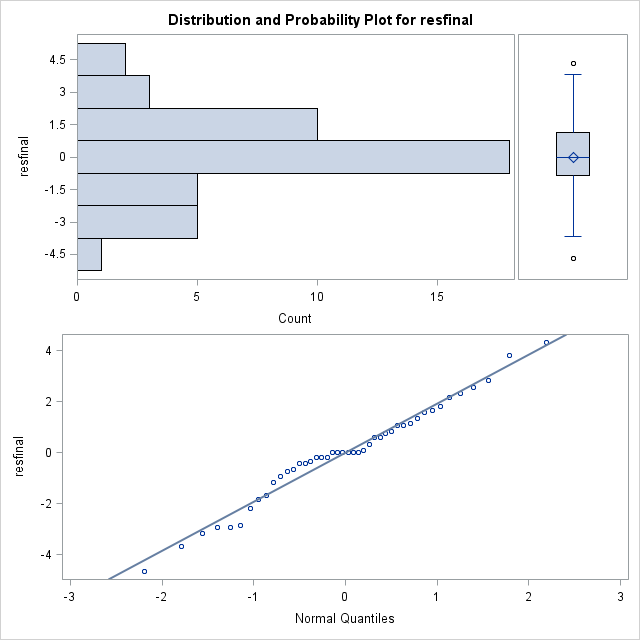


Figure 5. This is the probability plot. Since this plot is linear and the data points are close to the regression line, we can conclude that the data is normal.

Figure 6. The average days to germination of the tomato seeds with different soil and tomato feed treatments. The non-organic with feed lead to the shortest number of days to germination.

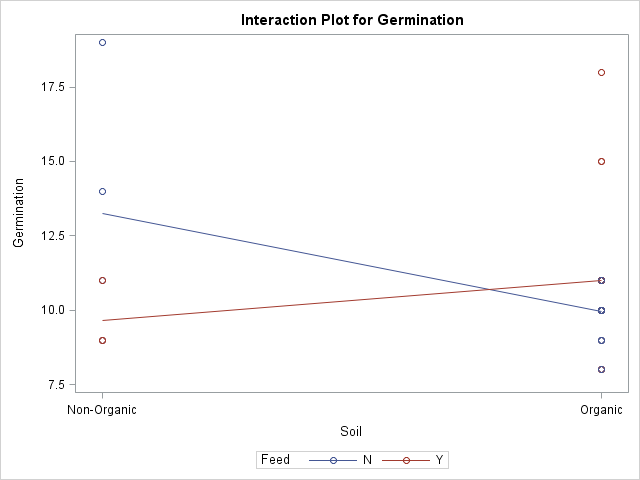


Figure 7. The interaction plot of soil type and the absence or presence of tomato feed. If the lines were parallel, there would be no interaction. Since the lines cross, we can tell that there is a significant interaction between soil type and tomato feed on days to germination in tomato seeds.

Table 1. The results of the Shapiro-Wilk (W) test for normality for the error variance (residuals). Since the test statistic is close to one and the p-value >.05, we fail to reject null hypothesis of normally distributed data for the transformed data set. For the untransformed data set, we can see that the p-value < .05 and therefore we reject the null hypothesis of normally distributed data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test for Normality | | | | |
| Shapiro-Wilk Test | Statistic | | p Value | |
| Untransformed Data | W | .907206 | Pr < W | .0024 |
| Transformed Data | W | .956874 | Pr < W | .1141 |

Table 2. The results of the Shapiro-Wilk (W) test for normality for the error variance (residuals). Since the test statistic is close to one and the p-value >.05, we fail to reject null hypothesis of normally distributed data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tests for Normality | | | | |
| Test | Statistic | | p Value | |
| Shapiro-Wilk | W | 0.979675 | Pr < W | 0.6207 |

Table 3. Complete ANOVA table. Results of the fixed effects factorial ANOVA of days between watering, absence or presence of tomato feed, soil type, and their interaction for days to germination of tomato seeds. Since p < .05 for only the soil feed interaction, we can conclude that it is the only treatment that is having a significant effect on the days until germination.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **Soil** | 1 | 0.05057799 | 0.05057799 | 1.87 | 0.1818 |
| **Days** | 2 | 0.12869643 | 0.06434821 | 2.37 | 0.1099 |
| **Soil\*Days** | 2 | 0.09654011 | 0.04827005 | 1.78 | 0.1854 |
| **Feed** | 1 | 0.00149074 | 0.00149074 | 0.05 | 0.8162 |
| **Soil\*Feed** | 1 | 0.36824000 | 0.36824000 | 13.58 | 0.0009 |
| **Days\*Feed** | 2 | 0.01164411 | 0.00582206 | 0.21 | 0.8080 |
| **Soil\*Days\*Feed** | 1 | 0.07340480 | 0.07340480 | 2.71 | 0.1100 |
| **Error** | 31 | 0.84053326 | 0.02711398 |  |  |
| **Corrected Total** | 41 | 1.57112743 |  |  |  |

Table 4. Complete ANOVA table. Results for the fixed effects ANOVA of the reduced model of soil type and absence or presence of tomato feed and their effect on days to germination of the tomato seeds. Since p > .05 for both treatments, we fail to reject the null hypothesis and assume that neither of these treatments is having an effect on days to germination.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **Soil** | 1 | 9.21904762 | 9.21904762 | 1.61 | 0.2119 |
| **Feed** | 1 | 0.91053864 | 0.91053864 | 0.16 | 0.6922 |
| **Error** | 39 | 223.2037471 | 5.7231730 |  |  |
| **Corrected Total** | 41 | 233.3333333 |  |  |  |

Table 5. Complete ANOVA table. Results of the fixed effects factorial ANOVA of days between watering, absence or presence of tomato feed, soil type, and their interaction for height of tomato seedlings. Since p > .05 for all of the treatments and their interactions, we can conclude that no treatment or interaction is having a significant effect on the seedling height.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **Soil** | 1 | 6.91414141 | 6.91414141 | 1.39 | 0.2474 |
| **Days** | 2 | 12.62178363 | 6.31089181 | 1.27 | 0.2954 |
| **Soil\*Days** | 2 | 9.85043860 | 4.92521930 | 0.99 | 0.3831 |
| **Feed** | 1 | 0.61602564 | 0.61602564 | 0.12 | 0.7274 |
| **Soil\*Feed** | 1 | 0.86250971 | 0.86250971 | 0.17 | 0.6801 |
| **Days\*Feed** | 2 | 13.32702020 | 6.66351010 | 1.34 | 0.2767 |
| **Soil\*Days\*Feed** | 2 | 20.81944444 | 10.40972222 | 2.09 | 0.1402 |
| **Error** | 32 | 159.3750000 | 4.9804687 |  |  |
| **Corrected Total** | 43 | 224.3863636 |  |  |  |

SAS CODE:

Germination Data

**PROC** **IMPORT**

DATAFILE="E:\final.xlsx"

OUT=WORK.final

REPLACE

DBMS=XLSX;

**run**;

**data** trfinal;

set final;

lngerm=log(germinati

sqgerm=(germination)\*(germination);

sqrtgerm= sqrt(germination);

lg2germ=log2(germination);

lg10germ=log10(germination);

**run**;

**proc** **glm** data=final;

class soil days feed;

model germination = soil|days|feed;

output out=resids r=resfinal p=prefinal;

**run**;

**Proc** **univariate** data=resids normal plot;

var resfinal;

**run**;

**Proc** **Plot** Data = resids;

Plot resfinal \* prefinal;

**run**;**proc** **glm** data= trfinal;

class soil days feed;

model lngerm = soil|days|feed;

output out=trresids r=trresfinal p=trprefinal;

**run**;

**quit**;

**Proc** **univariate** data=trresids normal plot;

var trresfinal;

**run**;

**Proc** **Plot** Data = trresids;

Plot trresfinal \* trprefinal;

**run**;

**Proc** **Means** data = final mean stderr;

class soil days feed;

var germination;

**run**;

Reduced Moodel:

**PROC** **IMPORT**

DATAFILE="E:\final.xlsx"

OUT=WORK.final

REPLACE

DBMS=XLSX;

**run**;

**data** trfinal;

set final;

lngerm=log(germination);

sqgerm=(germination)\*(germination);

sqrtgerm= sqrt(germination);

lg2germ=log2(germination);

lg10germ=log10(germination);

**run**;

**proc** **glm** data=final;

class soil feed;

model germination = soil feed;

output out=resids r=resfinal p=prefinal;

**run**;

**Proc** **univariate** data=resids normal plot;

var resfinal;

**run**;

**Proc** **Plot** Data = resids;

Plot resfinal \* prefinal;

**run**;**proc** **glm** data= trfinal;

class soil feed;

model lngerm = soil feed;

output out=trresids r=trresfinal p=trprefinal;

**run**;

**quit**;

**Proc** **univariate** data=trresids normal plot;

var trresfinal;

**run**;

**Proc** **Plot** Data = trresids;

Plot trresfinal \* trprefinal;

**run**;

**Proc** **Means** data = final mean stderr;

class soil feed;

Seedling Height:

**PROC** **IMPORT**

DATAFILE="E:\final.xlsx"

OUT=WORK.final

REPLACE

DBMS=XLSX;

**run**;

**proc** **glm** data=final;

class soil days feed;

model size = soil|days|feed;

output out=resids r=resfinal p=prefinal;

**run**;

**Proc** **univariate** data=resids normal plot;

var resfinal;

**run**;

**Proc** **Plot** Data = resids;

Plot resfinal \* prefinal;

**run**;

**Proc** **Means** data = final mean stderr;

class soil days feed;

var size;

**run**;