DEPARTMENT OF STATISTICS   THE UNIVERSITY OF BRITISH COLUMBIA

STAT 404 Term Project 2017

Judy Li Ming Ye	31774136
Xueyang Zhao	50197145
Yuwen Yu	20065141
Wenyue Li	48008149

1. Abstract
The purpose of our study is to find the best conditions for mung bean sprout growth using a half fraction of 2^4 factorial-design. We analyzed our data using different models and found whether each treatment effect is significant. When the treatment effects were suggested to be significant, we compared the different treatment levels using the treatment contrast effects and recommended prefered factor levels. Blocking factor was also added to our design, confounding it with our highest-order interaction effects. We concluded that warmer temperature and darker environment would be a better environment for bean sprout growth.

2.Introduction
Bean sprout, as a popular ingredient in our daily life, is a kind of vegetable which contains much vitamin C and dietary fiber. And the way bean sprout produces is relatively feasible for normal family. There are numerous kinds of bean sprout such as the mung bean sprout, soybean sprout and black bean sprout in the market, but unfortunately, not all of them have high quality. Therefore, it would be of interest to understand how the environmental conditions affect the sprout growth. It is true that the sprout growth may depend on several different factors, thus, we conducted an experiment that investigated the effect of factors including temperature, light, nutrients, and the kind of the bean in the sprout growth process. With some background research, we believe that the expected growth cycle of mung bean sprout is 2-6 days.

The rest of the report is as follows. **Details of the experimental design** will be given in Section 3,**Statistical analysis** will be given in Section 4, **Conclusion and discussion** will be given in Section 5. **Appendix** in Section 6 includes various tables and figures that are referenced throughout the report.

3. Experimental Design
3.1 The Experimental Plan
This experiment is a 2^{4-1} factorial design, which is a fractional factorial with 4 factors and 8 runs. *Eight* treatment combinations were constructed and randomized to the *eight* groups of inorganic mung beans; each group of inorganic beans contained 10 mung beans. Note that randomization decorrelates some possible unknown variables with the treatment factor of interest, allowing us to establish a causal relationship between the treatment factors and the response variable. Please refer to Table 1 and Table 2 on how we randomized them.

For each experimental run, we first randomized 10 inorganic mung beans in every 500 ml Nestle water bottle, then added 15 ml of water to the bottle. The same amount of water was then added to each bottle every day to avoid further possible blocking factors that may affect the response. At the end of the five-day period, the sprout length of each bean is measured and recorded, then the mean length of all 10 sprouts in one run is calculated.

3.2 Factors
Some factors we considered in our experiment are as follows.

Treatment Factors	Treatment Levels	Descriptions
Temperature (T)	Low vs. High	Low: $8 \circ C$ High: $25 \circ C$
Light (L)	Dark vs. Bright	Dark: Complete Darkness as the beans were completely covered up Bright: a 40 W, 120 V lamp shined directly

		on the beans
Nutrient (N)	Absence vs. Presence	Absence: No nutrient added Presence: 3 drops of "Jobes Drip Feeder"

- Blocking Factor: organicity of the mung beans: no vs. yes

3.3 The Response
- The mean length of the mung bean sprout in centimeters after 5 days

4. Statistical Analysis
4.1 Preliminary analysis
To avoid other random errors, we aimed to keep other factors constant. As stated in the **Experimental Design** section, each of our experimental run were kept constant with the same container and same amount of water. Each run was done by the same experimenter to avoid further variances between the results.

We decided to write the regression model for the 2-level experiment as represents

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_m + (\alpha\beta)_{ji} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + E_{ijk} \quad (i=1,2;\ j=1,2;\ k=1,2,\ m=1,2)$$

where Y_{ijk} is the mean length response for level i of *temperature,* level j of *light* and level k of *nutrients*; μ represents the *overall mean* effect; α_i represents the *temperature* treatment effect on the *mean* mean of Y. Similarly, β_j and γ_k are the effects of the *light* treatment factor and *nutrients* treatment factor, respectively. $(\alpha\beta)_{11}$,..., $(\alpha\beta)_{ji}$ are interaction effect of *temperature* and *light* on the *mean* response. Similarly $(\alpha\gamma)_{11}$,...,$(\alpha\gamma)_{ik}$ and $(\beta\gamma)_{11}$,...,$(\beta\gamma)_{jk}$ for the interaction effects of *temperature* and *nutrient* and of *light* and *nutrient* respectively. The random errors E_{ijk} are assumed to be independent $N(0, \sigma^2)$ random variables. The 3-factor interaction effect and all interactions involving the blocking factor are assumed negligible.

Since there are 3 treatment factors and 1 blocking factor in our experimental design, some insignificant effect needed to be dropped to leave at least one degree of freedom for residual.
The interaction plot in Figure 4.5.2 shows how the mean length of bean sprout depends on the levels of both temperature and light treatment factors. The mean response was plotted against the temperature for different levels of nutrients. It suggested that the two lines are relatively parallel and there weren't any interaction between the two factors. Similarly in Figure 4.5.3, the mean response was plotted against the temperature for different levels of light; the observed effect of temperature shows much the same pattern for each light level. So the interaction effects of temperature and nutrient and of temperature and light can be ignored here. Therefore, we resulted with the following model to work with for the rest of our report:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_m + \beta\gamma_{jk} + E_{ijk} \quad (i=1,2;\ j=1,2;\ k=1,2)$$

Moreover, please refer to the analysis section for further explanations.

Because of the little experimental runs, we used half-normal plot to identify the significant main and interaction effects, which is presented on Figure 4.1.1. In this section, we also assumed the 3-factor interaction effect to be negligible so it leaves some degrees of freedom for residuals. This allowed us to further examine the F-test and confidence intervals for our experiment of interest.

4.2 Analysis of Variance

In order to compare the significance of each treatment factor, we first consider the linear regression model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{23} x_2 x_3 + E$$

Where β_0 represents the *overall* effect, and β_1, β_2, β_3 are estimate coefficients for *temperature, light* and *nutrient,* respectively. β_4 is an estimate coefficient for *block.* In additions, x_1, x_2, x_3, and x_4 denote the categorial variables being either *0 or 1,* where 0 represents *low* level for *temperature*, *dark* level for *light*, and *absence* for *nutrient,* and *no* for *organicity* of the bean*;* the categorical value *1* represents the remaining level for each treatment factor and blocking factor. $E_1,...,E_8$ are assumed to be independent $N(0, \sigma^2)$ random variables.

Temperature:

First, we constructed the hypotheses:

$$H_0: \beta_1 = 0 \qquad H_a: \beta_1 \neq 0$$

According to output of the linear model (Please refer to Figure 4.2.1), we obtained an estimate coefficient of -5.8425cm for *temperature*. The standard error of *temperature* was found to be 0.03783 cm. Then we obtained a confidence interval of

$$\beta_1 \pm t_{2,\,0.975}\ se(\beta_1) = -5.8425 \pm 4.303 * 0.03783 = [-6.005, -5.679] \text{ cm}$$

Thus, we are 95% confident that the effect of changing temperature from 8 to 25 $\circ C$ is an increase in mean sprout length of [-6.005 , -5.679].

Brightness:

$$H_0: \beta_2 = 0 \qquad H_a: \beta_2 \neq 0$$

Consider the same linear model. Given the estimating coefficients β_2 and the standard error of β_2, we obtained a t-value of 8.691. This time we can compare the t-value with the $t_{2,0.975} = 4.303$. Since t-value = 8.691 is greater than $t_{2,0.975} = 4.303$, the *light* treatment factor is significant at 5% level.

Nutrient:

$$H_0: \beta_3 = 0 \qquad H_a: \beta_3 \neq 0$$

The estimating coefficient for *nutrient is* -0.185. The standard error of *nutrient* was found to be 0.0535. Then we obtained a confidence interval of

$$\beta_3 \pm t_{2,\,0.975}\ se(\beta_3) = -0.185 \pm 4.303 * 0.0535 = [-0.4152 , 0.04521] \text{ cm}$$

Additionally, ANOVA table also suggested that the *presence* versus *absence* of *nutrients* is not significant at the 5% level.

Organicity of mung beans

$$H_0: \beta_4 = 0 \qquad H_a: \beta_4 \neq 0$$

Consider the same linear model. Given the estimating coefficients β_4 and the standard error of β_4, we obtained a |t-value| of $\left|\frac{-0.10250}{0.03782}\right| = 2.709$. Since $t_{2,0.975} = 4.303$ and that |t-value| is smaller, the *organicity* blocking factor is not significant at 5% level.

Now looking at the interaction effect on the ANOVA table:

- The p-value of the interaction between *light* and *nutrient*s is 0.07789, almost significant at 5% level. However, according to the interaction plot between Light and Nutrients on Figure 4.5.1 it shows that the two lines are nonparallel to each other, suggesting that they do have little interaction effects. Further analysis of the interaction plot will be described in section 4.5.

In this selected model, we obtained 3 degrees of freedom for each main treatment factor effect, 1 degree of freedom for 2-factor interaction effect and 1 degree of freedom for block. Therefore, we are left with 2 degree of freedom for residual.

4.3 Model checking

Box-cox

In order to check whether we need to modify our data, we used box-cox to verify. As was seen on Figure 4.3.1, λ=1 is included in our 95% confidence interval of the box-cox plot, we do not need to transform our data.

Residuals plot

Figure 4.3.2 gives a scatterplot of the residual values (the difference between the observed and fitted values) against the fitted values. This plot shows that residuals are randomly located around the horizontal line 0 and have no pattern, which means the model fits the data well.

4.4 Contrast

Since the temperature was found to be very strongly significant, we will only compared and contrast the low versus high level of the temperature.

First, we set $\kappa_1 = \alpha_1 - \alpha_2$, where α_1 represents low temperature and α_2 represents warmer temperature. The hypotheses are as follows:

$$H_0:\ \kappa_1 = 0 \qquad\qquad H_a:\ \kappa_1 \neq 0$$

Since the *E[MS(Residuals)]* is an unbiased estimator for the variance, we used the same *MS(Residuals)* from the ANOVA table in section 4.2 to estimate the variance of the contrast. At the end, we obtained $k = -5.8425\ cm$ with $se(k) = 2 * \sqrt{\frac{MS(Residual)}{n}} = 0.05\ cm$, which led to the 95% confidence *interval of [-6.4778, -5.2072].* This suggested that warmer temperature is better for bean sprout growth.

Then, we set $\kappa_2 = \beta_1 - \beta_2$ where β_1 represents Dark and β_2 represents Bright, and κ_3 corresponds Nutrients effect. The hypotheses are similar to that of κ_1 . It is noted that the confidence interval of κ_2 and κ_3 include 0, which means there is no statistically significant difference between the two levels of both Light and Nutrients. And the contrast table is given in Table 3.

4.5 Recommended factor levels

From the interaction plot on Figure 4.5.1, the estimated mean length of mung bean sprouts decreases when nutrient changes from the level *presence* to *absence* at the light level *bright*. On the other hand, the estimated mean of mung bean sprouts length increases slightly when light was at the level *dark*. Thus, the plot shows an observed interaction effect between these two factors. However, either in the absence or the presence of nutrients, the mean length of the sprouts is longer in a *dark* condition.

As stated from section 4.4, the 95% confidence interval of *k* is *[-6.4778, -5.2072] cm,* which means colder temperature gives a shorter mean sprout length. In addition to that, the interaction plot in Figure 4.5.2 also suggested that in either level of the treatment factor *nutrient,* mean bean sprout lengths are longer in the warmer temperature. Therefore, we conclude that the mung bean sprouts better at warmer temperature.

The interaction plot in Figure 4.5.3 shows that no matter what level the nutrient is *absence* or *presence*, the warmer temperature always yields longer sprouts.

Therefore, we conclude that the condition of warmer temperature, darker environment is better for bean sprout growth.

4.6 Blocking Effects

Here, we set the B column of the contrast matrix to be confounded with the 3-factor interaction column of the contrast matrix. The generator is as follows:

$$B = 123$$

This is a *Revolution IV* design, where 2-interaction effects cannot be confounded with any main effects. By means, all the main effects and 2-factor interaction effects can be estimated separately from each other.

This alias also suggested that the block effect cannot be estimated separately from the *123* interaction effect. We could assume that the three interaction effect is negligible, and this assumption allows the ANOVA to estimate all 2-factor interaction effects. Note that we started with a model of only one 2-factor interaction effect, so only this one 2-factor interaction effect can be estimated as significant. Moreover, it can also be assumed that all 2-interaction effects involving B are unimportant.

5. Conclusion and discussion

In this study, two factors were found to have significant effects on the production of bean sprouts under home-made conditions, namely *Temperature* and *Light*. To be more specific, the warmer temperature $25 \circ C$ was significantly better for sprout growth than the cooler temperature $8 \circ C$ and the dark environment is also significantly better than the bright environment under artificial light. These analysis were done based on various techniques we've learnt from this course, including Randomized-Block Designs, Factorial Experiments, and Fractions of 2-level Factorial Designs etc,. However, there are still factors that we may need to consider and to improve in the future experiments.
If we were to do this experiment again, we might consider adding more blocking factors such as the "type of water" and the "amount of water" used. For instance, if the amount of water far exceeds the amount that the beans can absorb, the resulting beans might rot instead of germinate, which leads to zero growth on the sprouts. We might also change the treatment factor of artificial light to the natural sunlight, as we believed that natural lights are beneficial for most plant growth. As for the response variable, we could also take the length of the root hair into account, because the caulicle of the organic and inorganic mung beans have different smoothness. During the experimental procedure of planning and analyzing, we successfully applied what we have learnt from STAT 404 to the real-life experiment.

6. Appendix

	Effect			
Run	Temperature(+ as high, - as low)	Light (+ as bright, - as dark)	Nutrient (+ as presence, - as absence)	Organic = Temperature*Light*Nutrient (+ as organic, - as inorganic)
1	-	-	-	-
2	-	-	+	+
3	-	+	-	+
4	-	+	+	-
5	+	-	-	+
6	+	-	+	-
7	+	+	-	-
8	+	+	+	+

Table 1. Contrast matrix to estimate the effects for the experiment in two blocks of four runs

Block	Treatment factor			
	Run	Temperature	Light	Nutrient
Organic +	1 (4)	-	+	+
	2 (6)	+	-	+
	3 (7)	+	+	-
	4 (1)	-	-	-
Organic -	5 (3)	-	+	-
	6 (8)	+	+	+
	7 (2)	-	-	+
	8 (5)	+	-	-

Table 2. Random allocation of treatments to experimental units. (Run numbers in parentheses are from Table 1)

Contrast	$\widehat{k}$	se(κ)	Confidence interval

k_1 (temperature)	-5.8425	0.05	(-6.4778, -5.2072)
k_2 (light)	0.5925	0.05	(-0.0428, 1.228)
k_3 (nutrients)	0.0575	0.05	(-0.5778, 0.6928)

Table 3: Contrast table

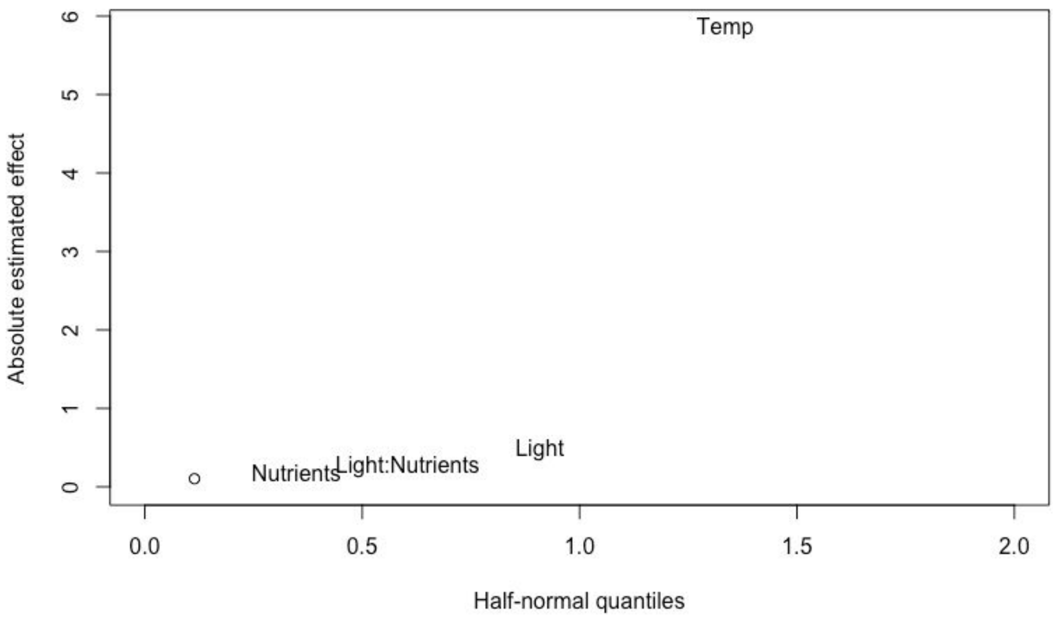


Figure 4.1.1: Half-normal plot

```
Call:
lm(formula = Y ~ Temp + Light + Nutrients + Organic + Light *
    Nutrients, data = dat)

Residuals:
      1       2       3       4       5       6       7       8 
-0.0050  0.0375 -0.0375  0.0050  0.0050 -0.0375  0.0375 -0.0050 

Coefficients:
                      Estimate Std. Error  t value Pr(>|t|)    
(Intercept)            6.50250    0.04633  140.339 5.08e-05 ***
Templow               -5.84250    0.03783 -154.433 4.19e-05 ***
Lightdark              0.46500    0.05350    8.691   0.0130 *  
Nutrientsyes          -0.18500    0.05350   -3.458   0.0744 .  
Organicyes            -0.10250    0.03783   -2.709   0.1135    
Lightdark:Nutrientsyes 0.25500    0.07566    3.370   0.0779 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0535 on 2 degrees of freedom
Multiple R-squared:  0.9999,    Adjusted R-squared:  0.9997 
F-statistic:  4823 on 5 and 2 DF,  p-value: 0.0002073
```

Figure 4.2.1 : lm output

```
                Df Sum Sq Mean Sq   F value   Pr(>F)
Temp             1  68.27   68.27 23849.646 4.19e-05 ***
Light            1   0.70    0.70   245.279  0.00405 **
Nutrients        1   0.01    0.01     2.310  0.26790
Organic          1   0.02    0.02     7.341  0.11350
Light:Nutrients  1   0.03    0.03    11.358  0.07789 .
Residuals        2   0.01    0.00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.2.2: aov output

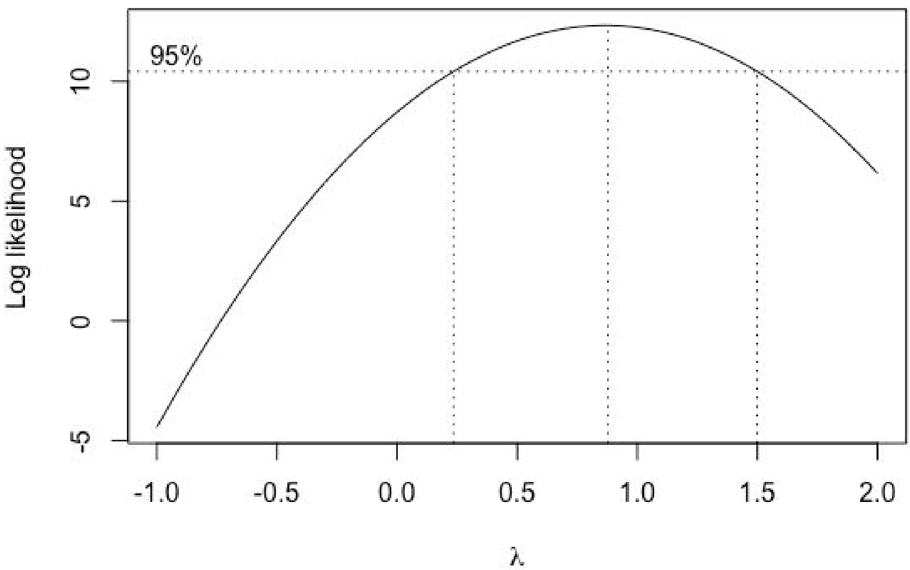


Figure 4.3.1: The Box-cox Analysis

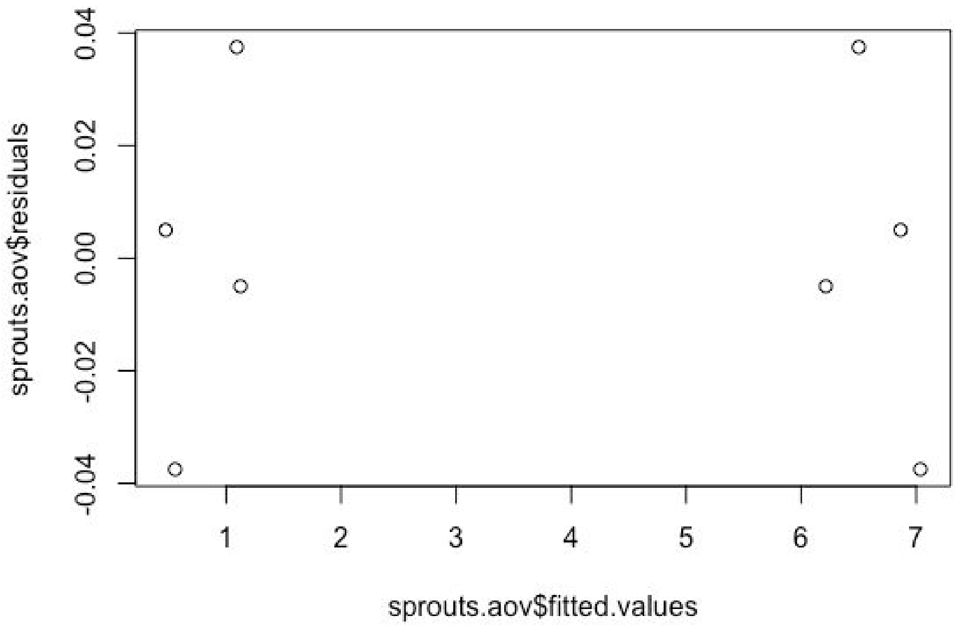


Figure 4.3.2: residuals versus fitted values

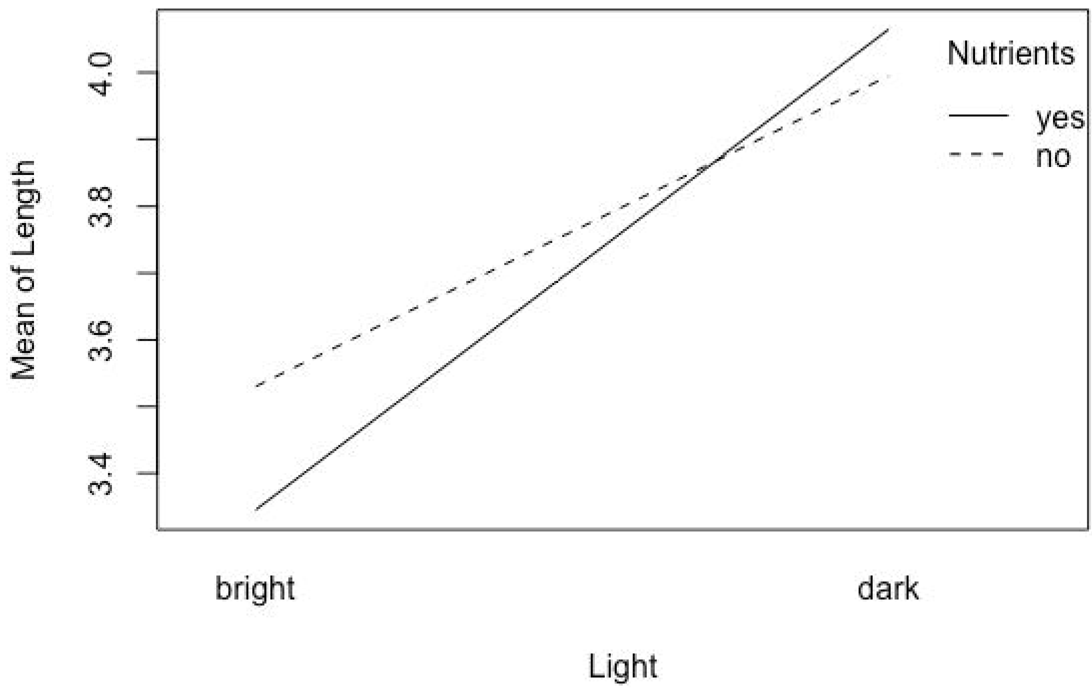


Figure 4.5.1 plot of interaction-effect of Brightness and Nutrients (Y-axis in cm)

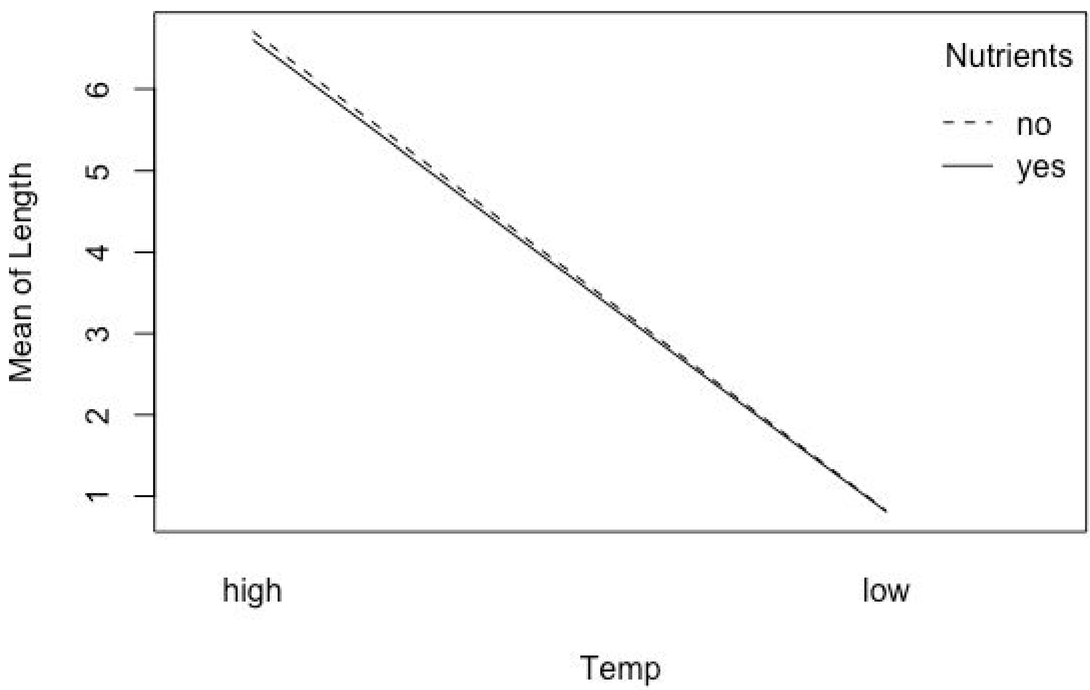


Figure 4.5.2 plot of interaction-effect of Temperature and Nutrients (Y-axis in cm)

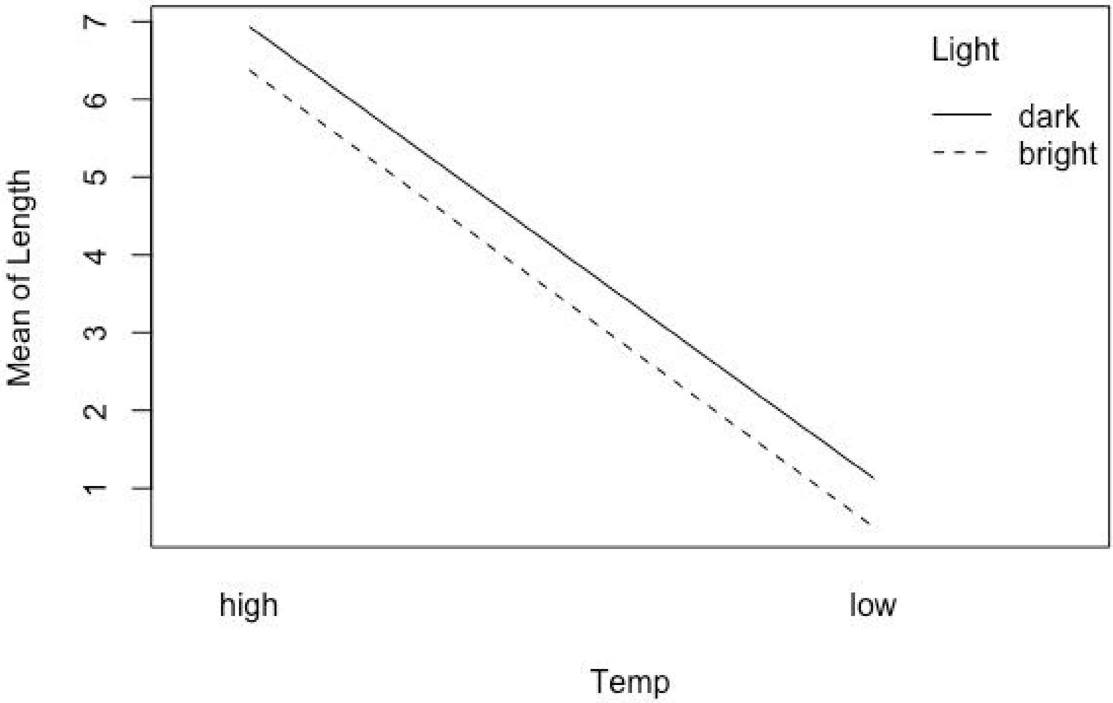
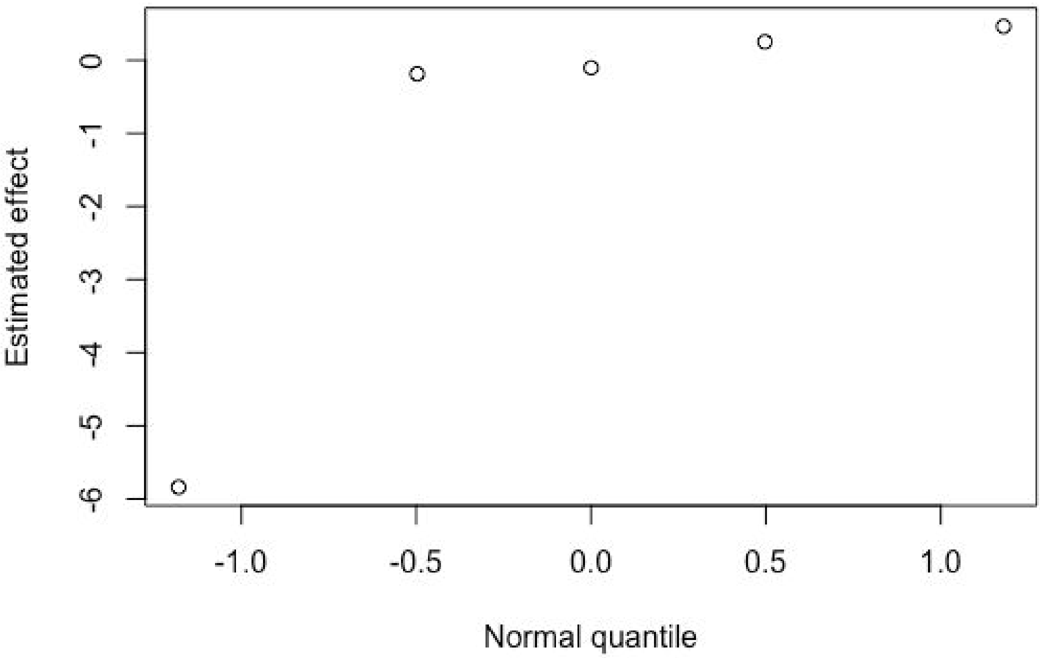


Figure 4.5.3 plot of interaction-effect of Temperature and Brightness (Y-axis in cm)



Normal Probability plot

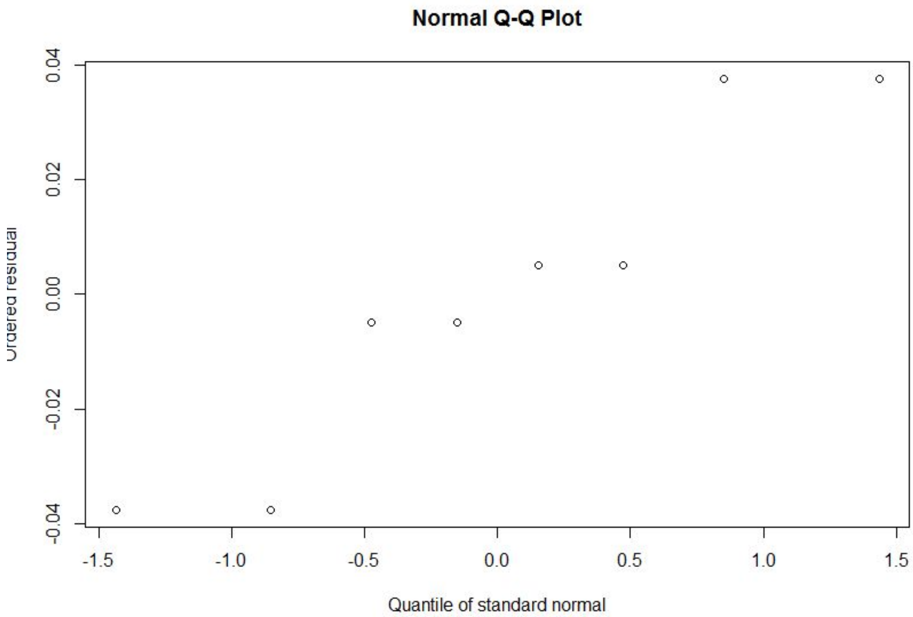


Figure: Normal Q-Q Plot

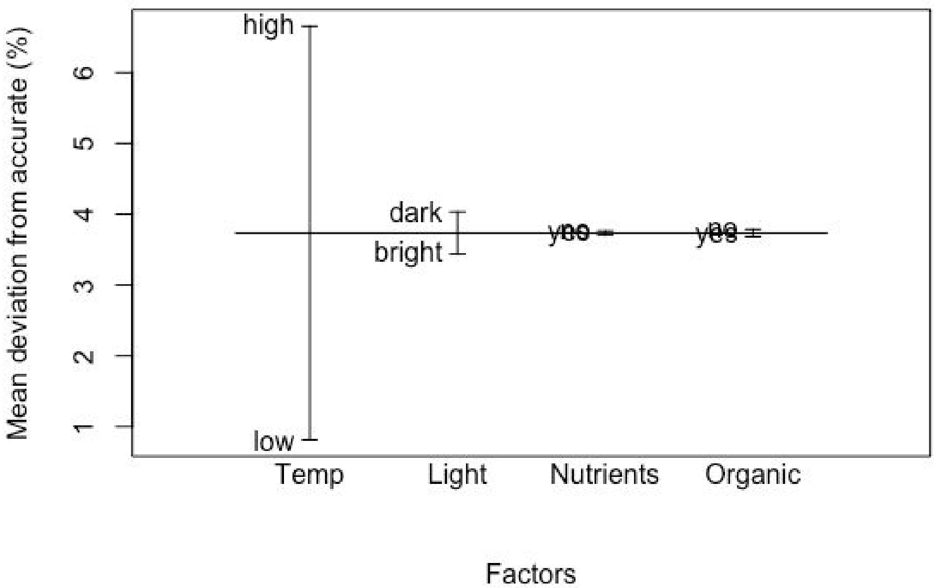


Figure: design plot