Investigating the Effect of Humidity on the Growth of Two Varieties of C.sativa

STAT 403

Ziying(301417046) Peng
Kaiwen(301354411) Gu
Shiqing(301385518) Liu

Part 1. Preamble

1.1 Purpose of the study

The purpose of this study was to explore how humidity affects the growth of two types (B52 and Northern Lights) of C. sativa plants. The objective was to determine if it is economically feasible to cultivate these plants in a high humidity greenhouse. Humidity levels are known to affect the growth and yield of these plants, and optimal growth tends to occur under high humidity conditions. However, maintaining such an environment in a large-scale production setting can be expensive, raising questions about the economic viability of this approach.

To investigate this issue, we will conduct experiments using up to 16 mini-greenhouses, each housing up to four pots with two seedlings each. The experimental design will be flexible in order to optimize resource use within budgetary constraints while ensuring thorough investigation of the effects of humidity on plant growth and yield.

1.2 The three basic structures

1.2.1 Treatment structure

The treatment structure for this experiment is based on a two-factor design with a factorial treatment structure, focusing on the effects of humidity levels and plant varieties on the growth and yield of C. sativa. The two factors are:
Humidity: This factor has two levels, representing low and high humidity conditions for plant growth.

1. Plant Variety: This factor includes two levels, representing the two varieties of C. sativa, B52 and Northern Light.
2. In this factorial treatment structure, all possible combinations of humidity levels and plant varieties are considered, resulting in a total of 4 treatment combinations (2 humidity levels x 2 plant varieties).

As the treatment structure is factorial, the statistical model used for analyzing the experimental data will include terms corresponding to the main effects of both factors, as well as their interaction. Specifically, the model will have the following terms:

1. Main effect of humidity (denoted as humidity).
2. Main effect of plant variety (denoted as plant).

Interaction effect between humidity and plant variety (denoted as humidity*plant).
By including these terms in the model, we can assess the individual and combined effects of humidity levels and plant varieties on the growth and yield of C. sativa plants.

1.2.2 Experimental Unit structure

This experiment utilizes a split-plot method with a two-level experimental unit structure. The main plot experimental unit is the greenhouse, with each greenhouse assigned to a specific humidity level. The subplot factor is the individual plant within each pot, irrespective of the plant variety. Each greenhouse is uniquely labeled and assigned to a humidity level (low or high), while each pot contains two seedlings randomly assigned to either B52 or Northern Light varieties. The observational unit is the single plant in each pot.

The experimental unit structure accurately reflects the hierarchical organization of the experiment, with greenhouses serving as the main plot experimental units and individual plants within pots acting as subplot experimental units. The statistical model will account for the greenhouse term (main plot experimental unit) and the residual term (subplot experimental unit) to analyze the effects of humidity levels and plant varieties on the growth and yield of C. sativa plants.

1.2.3 Randomization structure

The randomization structure for this split-plot greenhouse experiment consists of two separate levels of randomization.

Upper-level randomization (main plot): The main plot treatments, which are the humidity levels (low and high), are randomly assigned to the 16 greenhouses. This randomization can be done as a completely randomized design (CRD). Since we do not concern the other factor, therefore we do not use randomized complete block design (RCBD) to control other whether additional blocking factors. In this case, 8 greenhouses will be assigned to low humidity, and 8 greenhouses will be assigned to high humidity.

Lower-level randomization (subplot): Within each greenhouse (main plot), the subplot treatments, which are the plant varieties (B52 and Northern Light), are randomly assigned to the 4 pots. This randomization is done independently for each main plot.

In summary, the randomization structure for this split-plot greenhouse experiment includes random assignment of humidity levels to greenhouses (main plot) and random assignment of plant varieties to pots within each greenhouse (subplot). This randomization approach ensures that the effects of humidity levels and plant varieties on the growth and yield of C. sativa plants can be properly assessed and analyzed.

1.3 Design considerations

A split-plot design is a natural choice for experiments in which the experimental factors can be divided into two groups, one group of factors that are relatively easy to change or control, and the other group of factors that are more difficult to change or control. It allows for greater control and precision and incorporates inherent variability, resulting in a more accurate representation of real-world conditions compared to a completely randomized design.

Using a split-plot design allows the experimenter to focus on the effects of the main plot factor (humidity level) and the subplot factor (plant species) separately, and to study the interaction between these two factors. Repeating treatments independently by randomly assigning greenhouses to different humidity levels and pots to different plant species helped reduce bias and improve the precision of the estimates.

In contrast, a completely randomized design (CRD) may not be the best choice for this experiment, as it does not account for potential interaction effects between moisture levels and plant species, and may require a larger sample size to achieve the same level Precision in estimating treatment effects. CRD is also less efficient because it does not take advantage of the natural grouping of experimental units into greenhouse and potting hierarchies.

Part 2. Pre-test

2.1 Pre-test design

A pre-test was conducted using 8 greenhouses. The humidity levels (Low and High) were randomly assigned to the greenhouses. Within each greenhouse, 4 pots were randomly placed, and in each pot, 2 seedlings were randomly planted.

2.2 Variability estimates

After running the code and fitting the linear mixed-effects model, the estimated variability components were as follows:

Greenhouse standard deviation: 1.33795
Residual standard deviation: 0.43808

2.3 Required sample size

Using the obtained variability estimates, a power analysis was performed to determine the required sample size for detecting a biologically important difference. The following parameters were used for the power analysis:

Effect size: 0.1
Significance level: 0.05
Statistical power: 0.8
Comined standard deviation: 1.34

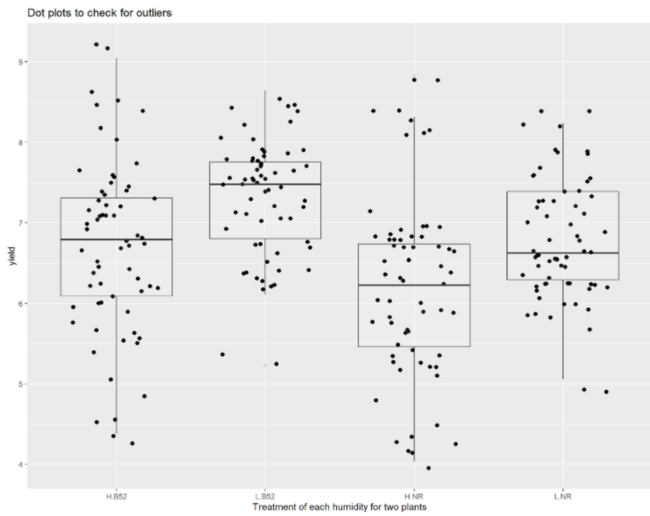
Based on the power analysis, the required sample size to detect a 10% difference in yield with a significance level of 0.05 and 80% power is 875.194. However, due to budget limitation, we are unable to gather 876 observation units.

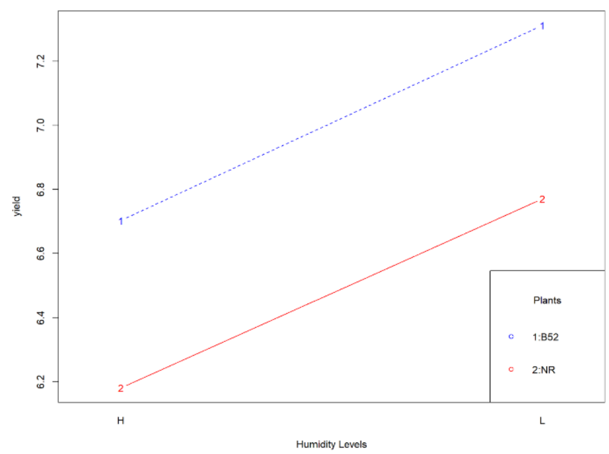
Part 3. Full experiment

3.1 Preliminary analysis and checking assumptions

Here we use *ggplot()* function to produce the side-by-side dot-plot. We also added random noise to the points in the plot to prevent overlapping of points that have the same or similar coordinates. From the box-plot below, there are two outliers or unusual points observed in Low Humidity with B52.

Here we use *ggplot()* function to produce the profile (interaction) plot. From the interaction plot below, we do not seem to indicate the presence of any large interaction since the lines are approximately parallel. The B52 appears to provide a higher mean yield.





Here we use *ddply()* function to produce the table of mean and standard deviations. From the table below, we can conclude that the standard deviations are approximate equal in each group.

	Trt	n.Trt	mean.Trt	sd.Trt
1	H.B52	32	6.702567	1.0979070
2	L.B52	32	7.311452	0.7363940
3	H.NR	32	6.181217	1.1950993
4	L.NR	32	6.770856	0.7476976

To conclude, satisfy the necessary assumptions of the ANOVA.

3.2 Statistical model

The statistical model is:

$$Yield = Humidity + Plant + Humidity * Plant + \varepsilon$$

The terms *Humidity*, *Plant*, and *Humidity*Plant* represents the treatment structure - a two-factor, complete factorial experiment.

The term *Error* is an experimental unit effect and specified as a random effect. Since the observational unit is the same as the sub-plot experimental unit, it is not necessary to specify it in the model.

H_0 : There is an interaction between plant variety and humidity.

H_a : There is no interaction between plant variety and humidity.

From the ANOVA test output below, since p-value less than 0.05. We can reject the Null hypothesis; we can conclude that the interaction between plant variety and humidity is not significant. For further study, we still need to check the main effects of Humidity and Plant.

```
Type III Analysis of Variance Table with Satterthwaite's method
                Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)    
humidity       0.07187 0.07187     1     6  0.3734 0.5635717    
plant          3.01324 3.01324     1    54 15.6547 0.0002235 ***
humidity:plant 0.00986 0.00986     1    54  0.0512 0.8218314    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

H_0 : The coefficients of Humidity in the model are all equal to zero.

H_a : At least one of the coefficients of Humidity is not equal to zero.

AND

H_0 : The coefficients of Plant variety in the model are all equal to zero.

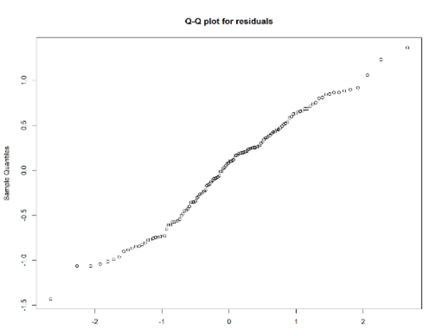
H_a : At least one of the coefficients of Plant variety is not *equal* to zero.

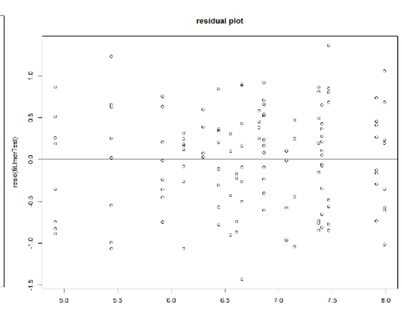
For the ANOVA test output below, we can observe that the p-value of Humidity larger than 0.05, we can reject the Null hypothesis, therefore, we conclude that the effect of Humidity factory is not significant. Besides, we can observe that the Plant variety has a significant effect on Yield.

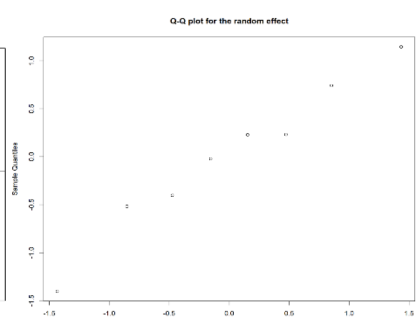
```
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)    
humidity 0.9081  0.9081     1 121.76  2.4664   0.1189    
plant    9.0218  9.0218     1 118.12 24.5040 2.49e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3.3 Residual check

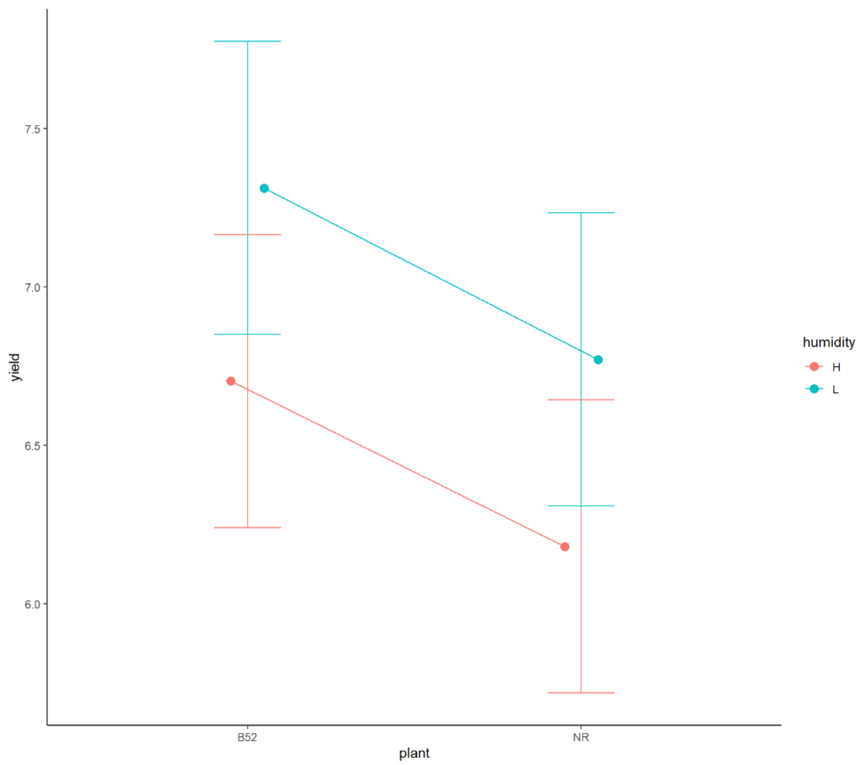
From the Q-Q plot for residuals below, we can observe that there are some points that depart the line in the two sides of the plot. For the Q-Q plot for the random subject effect and the residual plot below, we didn’t observe any major problems. Although there is some evidence of non-normality in the residuals, it is not serious. The residuals follow roughly normality.







The summary plot below shows the mean yield and the variability (represented by error bars) for different plant varieties under different levels of humidity. Each point represents the mean yield for a specific plant variety under a specific humidity level. The color of the point represents the humidity level. The error bars show the variability of the yield for each plant variety and humidity level. From the se bar below, we can conclude that there is not significant difference between yield in high humidity and low humidity environment.



Part 4. Conclusions

Based on the statistical analysis of the data using the model Yield = Humidity + Variety + Humidity*Variety + Error, it can be concluded that there is no significant main effect of humidity on the growth and yield of C. sativa plants. Additionally, the interaction effect between humidity and plant variety on the growth and yield of C. sativa plants is not significant. Therefore, these plants in full-scale production in high humidity level greenhouses is not economically viable.

From the experiment report above, we can observe that the main effect of plant variety was significant, indicating that the type of plant variety used does have an impact on the growth and yield of C. sativa. Therefore, we recommend that growers consider cultivating the B52 variety under low humidity conditions, as our analysis suggests that this combination is associated with the highest mean yield.