Length of home-grown mung bean sprouts: relation to temperature, oxygen and watering

20120809, 20048598, 20045458

February 23rd, 2021

Summary

Background: Recently, people are asked to stay at home as coronavirus cases are rising rapidly across England. If we were to grow vegetables like bean sprouts at home, we could reduce our shopping times for necessities thus helping protect the NHS. This study aims to find the optimal temperature, oxygen and watering for the growth of mung bean sprouts and explore the effects of these treatments.

Methods: An orthogonal designed experiment is conducted to explore the effects of three treatments separately and estimate the optimal range, followed by a full factorial designed experiment to further analyze the effects of treatments and interactions and search for the optimal treatment combination.

Results: The best condition for the growth of mung bean sprouts is mid-level temperature, low-level water and high-level oxygen. There exist interaction effects among temperature, water and oxygen.

Conclusion: When growing mung bean sprouts at home, we should keep them in a warm (about 25^oC), well ventilated places and water them properly (about 6 bottle caps or 10ml per day).

1 Introduction

COVID-19 has become a severe threat to global health and it is spreading rapidly across England. To help slow down the transmission of the virus and protect the NHS, we are not supposed to leave home except where necessary. By growing bean sprouts at home, we could reduce our shopping times for fresh vegetables. Thus, we are wondering what the best conditions are for home-grown mung bean sprouts. In the first place we hope to find the answers by search the Internet but it returned millions of results suggesting different conditions. Therefore, we decided to design an experiment to find the optimal conditions for the growth of mung bean sprouts and also explore the associations among the three treatments of interest: temperature, oxygen and watering.

2 Methodology and Data

2.1 Orthogonal Design

The orthogonal experimental design is a design method to study multi factors and multi-levels. It selects some representative points from the comprehensive test according to the orthogonality. The combinations selected by orthogonal design has the following two characteristics: (1) the entries in each column appear equally often; (2) the entry-combinations in any two columns appear equally often.

2.1.1 Pilot Study

Considering the variable effects that may affect the plant growth and within the experimental capacity, the amount of watering per day, the growth environment temperature and oxygen concentration are finally selected. According to some literature, the suitable temperature for plant growth is generally between 5 ^oC and 40 ^oC, an experiment is set every 5 ^oC. It is found that the suitable temperature range for bean sprout growth is 10 ^oC to 30 ^oC, and most of the bean sprouts will die outside this range. The bottom area of the container for cultivating bean sprouts has a certain influence on the daily watering amount, but when the amount of water completely exceeds the beans, all the plants will die. Therefore, the appropriate amount of water in this experiment is 2 to 12 bottle caps (per day). There are also some other potential factors, such as light, which are also involved in the pilot study, and the best situation is selected in the formal experiment to ensure that the growth environment of bean sprouts is as good as possible, so as to avoid a large number of deaths or stop growth, affecting the results.

2.1.2 Experiment

An orthogonal design table $L_9(3^3)$ is used for design. More specifically, water quantity as factor A with three levels of 2, 4 and 6 bottle caps; the temperature of growing environment as factor B with three levels of placed in the window (15 ^oC), placed at room temperature (20 ^oC) and placed on the heater (25^oC); oxygen concentration as factor C with three levels of the top of the bottle is completely covered (Low), half covered (Median), and completely uncovered (High). Firstly, we soaked a large number of mung beans in warm water for germination treatment. After 24 hours, we selected the beans that successfully sprouted without damage to ensure the consistent growth environment of beans before grouping. Then according to the orthogonal design, 40 beans were placed in 9 groups, and the sprout length was measured after 96 hours.

2.2 Factorial Design

A further completely cross-classified designed factorial experiment is conducted, using the same factors as above but with different levels. Now the three levels of water quantity are 6, 8 and 10 bottle caps. Those of temperature are on the balcony (10 ^oC), at room temperature with the heated turned up (25 ^oC), and on the heater (30

^oC). The levels of oxygen concentration remain unchanged. All levels are labeled 'low', 'medium', 'high' in order. Then a simple linear regression is applied to explore the significance of main effects and interaction effects, where a best subset selection is used to simplify the model.

3 Analysis and Results

3.1 Orthogonal Design with Repetition without Interactions

The aforementioned orthogonal design table method is applied to analyze the fractional factorial design in the section, which is straightforward as a result of ignorance of interactions. Firstly, record and summarize some fundamental statistics information of the experiment into the design table (shown in Table 1). More specifically, the top half of the table records 9 groups of experimental conditions and corresponding 40 experimental results (growth length of bean sprouts) and summarizes the total growth length of bean sprouts under each experimental condition (Total). The bottom half of the table summarizes the total (T_i) and average (m_i) output at each level, and the largest differences (Range) of average output at each level for each factor. Moreover, the range of average sprout length indicates the importance rank of factors should be $A > C > B$. The means plot (shown in Figure 1) can be got according to the average sprout length (m_1, m_2, m_3) at different levels of each factor, which can be used to observe its growth tendency.

Table 1: Design table and data collection

No	A	B	C	40 Outputs (cm)	Total (cm)
1	2(1)	15(1)	High(3)	4.0, 3.9, 4.2, 4.7, ..., 2.7, 1.4, 2.0, 2.6	110
2	2(1)	20(2)	Median(2)	5.4, 5.6, 7.0, 7.2, ..., 2.3, 1.5, 2.1, 1.6	168
3	2(1)	25(3)	Low(1)	1.6, 1.7, 1.3, 0.5, ..., 0.4, 1.8, 0.3, 0.5	34.4
4	4(2)	15(1)	Median(2)	2.1, 4.8, 3.0, 3.0, ..., 3.5, 4.2, 1.6, 2.8	137.3
5	4(2)	20(2)	Low(1)	2.1, 2.9, 2.6, 4.8, ..., 2.2, 1.2, 1.4, 0.8	101.7
6	4(2)	25(3)	High(3)	5.9, 8.8, 9.0, 10.8, ..., 3.8, 3.4, 2.6, 2.7	296.3
7	6(3)	15(1)	Low(1)	13.0, 7.8, 11.4, 4.0, ..., 2.1, 1.0, 0.9, 0.9	207.5
8	6(3)	20(2)	High(3)	8.3, 7.2, 6.4, 5.8, ..., 4.2, 4.7, 2.5, 2.8	222.6
9	6(3)	25(3)	Median(2)	11.0, 11.1, 9.9, 10.0, ..., 1.8, 1.6, 1.4, 1.7	259.1
T_1	312.4	454.8	343.6		
T_2	535.3	492.3	564.4		
T_3	689.2	589.8	628.9		
m_1	2.6033	3.7900	2.8633		
m_2	4.4608	4.1025	4.7033		
m_3	5.7433	4.9150	5.2408		
Range	3.1400	1.1250	2.3775		

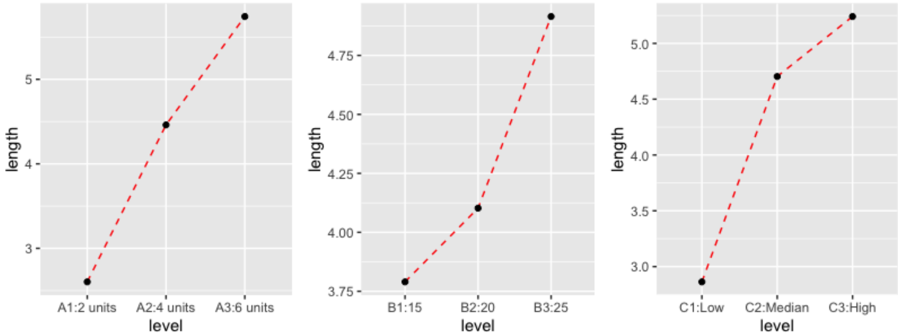

Figure 1: Comparison of length at different levels

Secondly, construct the statistical model for the experiment as $y_n = \mu + a_i + b_j + c_k + \epsilon_n$. Let a_i be the main effect of factor A at i-th level, b_j be the main effect of factor B at j-th level, c_k be the main effect of factor C at k-th level, where $i, j, k = 1, 2, 3$. Under the condition of $\sum_{i=1}^{3} a_i = \sum_{j=1}^{3} b_j = \sum_{k=1}^{3} c_k = 0$, the overall mean of y is calculated by $\hat{\mu} = \bar{y} = \frac{1}{360} \sum_{n=1}^{360} y_n = 4.2692$.

Next, estimate all the main effects by $\hat{a}_i = m_{i1} - \hat{\mu}$, $\hat{b}_j = m_{j2} - \hat{\mu}$, $\hat{c}_k = m_{k3} - \hat{\mu}$. Therefore,

$$\hat{a}_1 = -1.6659, \hat{a}_2 = 0.1916, \hat{a}_3 = 1.4741,$$
$$\hat{b}_1 = -0.4792, \hat{b}_2 = -0.1667, \hat{b}_3 = 0.6458,$$
$$\hat{c}_1 = -1.4059, \hat{c}_2 = 0.4341, \hat{c}_3 = 0.9716.$$

Due to the experiment purpose of finding the conditions to increase the growth length of sprouts, $A_3B_3C_3$ seems to have the 'best' value base on the estimated main effect, which is consistent with the means plot. Thus, the predictor of y at the 'best' level combination should be

$$\hat{y} = \hat{\mu} + \hat{a}_3 + \hat{b}_3 + \hat{c}_3 = 4.2692 + 1.4741 + 0.6458 + 0.9716 = 7.3607.$$

Finally, construct the ANOVA table and apply hypothesis testing to judge which factors have significant effects on the responses y at $\alpha = 0.01$.

Hypothesis 1: $H_o : a_1 = a_2 = a_3 = 0$ vs $H_1 : a_i \neq 0$, for some i.
Hypothesis 2: $H_o : b_1 = b_2 = b_3 = 0$ vs $H_1 : b_j \neq 0$, for some j.
Hypothesis 3: $H_o : c_1 = c_2 = c_3 = 0$ vs $H_1 : c_k \neq 0$, for some k.

Table 2: Anova table

Source	df	SS	MS	F	C.V. at 0.01
Water: A	2	598.1885	299.0943	57.86865	4.6658
Temperature: B	2	80.9375	40.46875	7.82988	4.6658
Oxygen: C	2	373.0805	186.5403	36.09174	4.6658
Error	353	1824.481	5.168502		
Total	359	2876.688			

According to the results from R code (shown in Table 2), $F_A > F_{2,353}$, $F_B > F_{2,353}$, and $F_C > F_{2,353}$, it implies that there are strong evidences to reject the null hypothesis. Hence, water, temperature, and oxygen factors have significant effects on the responses at $\alpha = 0.01$.

As stated above (see figure 1), the best combination is suggested to be $A_3B_3C_3$ as when the amount of water, temperature and oxygen reach level 3, the average length of bean sprouts is the largest. Therefore, it can be concluded that the case of six bottle caps of water, 25 degrees centigrade and with no cover is the best combination in this experiment. However, Figure 1 does not show inflection points. In other words, when amount of water and temperature (as oxygen concentration cannot be increased more) continue to rise, it is impossible to judge whether the length of bean sprouts is positively correlated through the existing experiments. At the same time, in order to simplify the experimental process and reduce the number of experiments, it is unable for orthogonal experiment to calculate the effect of two-way interaction terms. Based on these two reasons, some changes have taken place in the level selection of factorial experiments. First, because the higher oxygen content is above the test capacity, the oxygen level in factorial experiment is consistent with that before. In addition, according to the range of water and temperature suitable for bean sprout growth, the upper limit of temperature was increased to 30 degrees centigrade, and the upper limit of water was increased to 10 units.

3.2 Factorial Experiment

Following the above experiment, a factorial experiment including the best scenario found is conducted to take into account two-way interactions. To easily assess and remove the interactions, the factors with 3 levels need to be converted into dummy variables. Then a best subset selection is performed using all available dummy variables and interactions with six main effects forced in. The performance of the models selected is comprehensively evaluated by Adjusted R^2, Mallow's Cp (equivalent to AIC given normality assumption) and BIC. The summary of evaluation is as below:

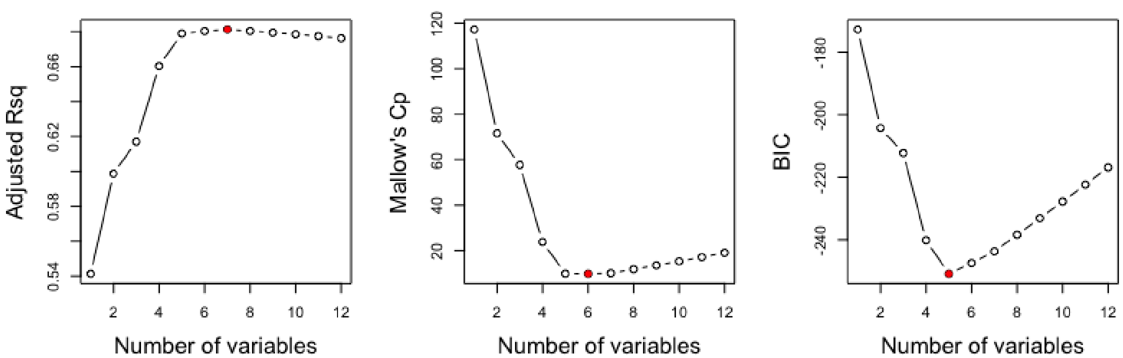

Figure 2: Best subset selection summary for the sprout data

From Figure 2, adjusted R^2 suggests M_7, whilst Mallow's Cp and BIC suggest M_6 and M_5, respectively. However, in the plot for Adjusted R^2, there appears to be little difference between models M_5,...,M_8, whilst the plot for Mallow's Cp statistic suggests there is little difference between models M_5, M_6 and M_7. Therefore we might regard the best model as M_5, which includes 5 interaction terms. The following Table 3 shows the coefficients and corresponding p-values for model M_5 (where T stands for temperature, O stands for oxygen, W stands for water, l stands for 'low', m stands for 'medium', h stands for 'high'):

Table 3: Coefficients for the Linear Model

Variable	(Intercept)	T_m	T_h	O_m	O_h	W_m
Coefficients	2.37	9.67	1.99	-0.42	0.83	-0.82
P-value	3.13e-05	6.48e-46	1.13e-02	4.48e-01	1.78e-01	6.88e-02

Variable	W_h	$T_m : W_h$	$T_h : O_m$	$T_h : O_h$	$T_h : W_h$	$O_h : W_h$
Coefficients	0.39	-6.04	7.09	6.62	-5.9	-3.31
P-value	6.11e-01	1.09e-09	1.58e-12	3.26e-11	2.39e-09	8.02e-05

From Table 3 we conclude that:

- The intercept is significantly different from zero with coefficients being 2.37, which means if a mung was grown under low-level temperature, oxygen and water treatment, the length would be expected to be 2.37 cm.
- The coefficient for T_m is 9.67 with p-value being 6.48e-46, which means if no interaction terms was taken into consideration (not under high-level water treatment in this case as we include $T_m : W_h$ in the model), the length of a mung grown under mid-level temperature would be expected to be 9.67 cm longer than that under low-level temperature.
- The interpretation of the coefficients for interactions is complicated. For example, the coefficient for T_m:W_h is -6.04, which seems to indicate that if treated with mid-level temperature and high-level water, a mung sprout would be expected to be 6.04 cm shorter than any other cases. However, we know that is not true as we need to consider other potential interaction effects. To give a general idea of the effects all together, the table of predicted length for 27 ($3 \times 3 \times 3$) distinct cases in total is shown below:

Table 4: Predicted sprout length for a total of 27 cases

$\mathbf{T_l}$	W_l	W_m	W_h	$\mathbf{T_m}$	W_l	W_m	W_h	$\mathbf{T_h}$	W_l	W_m	W_h
O_l	2.37	1.55	2.76	O_l	12.04	11.22	6.39	O_l	4.36	3.53	-1.16
O_m	1.95	1.13	2.34	O_m	11.62	10.80	5.97	O_m	11.03	10.21	5.51
O_h	3.20	2.38	0.28	O_h	**12.87**	12.05	3.91	O_h	11.81	10.98	2.98

It is clear that under mid-level temperature (25^oC), low-level water (6 bottle caps) and high-level oxygen (uncovered) treatment, the expected length reaches the highest point at 12.87 cm. It coincides with the best combination suggested in the previous orthogonal designed experiment.

4 Discussion

Several limitations of this study should be acknowledged. First, due to the small sample size, we have to set relatively great differences between the treatments of different groups to make the treatment effect significant. If a more precise range of the treatments was desired, a larger sample size would be required. Second, there is a negative value in Table 4 which is impossible in reality. The model might exaggerate the negative effects of high-level temperature, low-level oxygen, along with high-level water. To get a more reasonable prediction, we may need a larger data set or try to fit a different model.

References

[1] Cornell, J.A., Hamada, M.S. (2002) 'Experiments with Mixtures', 3nd Ed., *Wiley.*

[2] Fang, K.T., Li, R. and Sudjianto, A. (2005) 'Design and Modeling for Computer Experiments', *Chapman & Hall/CRC Press*, London.

[3] Montgomery, D.C. (2007) 'Design and Analysis - Analysis of Experiments', 6th Ed., *Wiley.*

[4] Wu, C.F., Hamada, M.S. (2000) 'Experiments: Planning, Analysis, and Parameter Design Optimization', *Wiley.*

Word Count: 1817