Examine the efficacy of ascaroside #18 in controling Escherichia Coli on alfalfa and fenugreek seeds and sprouts

Xueyan Hu

2024-02-23

# 1. Summary/Abstract

In this project, I will be concentrating on analyzing my own data gotten from the experiment that I carried on by my own. And I will be trying to figure out if there will be any relation between interested variables and observations.

# 2. Introduction

## 2.1 General Background Information

The core idea of this research comes from the following facts. First, sprouts are usually considered nutritious but they are consumed in raw mainly for foods like salad and sandwiches. Therefore, there have been a lot of foodborne pathogen outbreaks occurred across the states in the decades (Miyahira & Antunes, 2021), which means that there is an urgent need to reduce the contamination of sprouts with foodborne pathogens.

At the meantime, a recently study has revealed that a newly derived chemical could enhence the immune resistance of some certain plants to the pathogens including bacteria and virus but not by killing them directly(Manosalva et al., 2015).

Based on the facts above, a hypothesis is raised that if this chemical can also be applied on seedling seeds to control the level of contamination during sprouting process.

## 2.2 Description of data and data source

The data is obtained from experiment that I did it by myself, which has 160 observations and several variables. The data is the result of bacterial populations of foodborne pathogens on sprouts that are collected from different types of plants at different germinating time points under different treatments for the seeds

## 2.3 Questions/Hypotheses to be addressed

I would like to figure out the effect of different treatments on controlling bacterial populations as well as the factors also may be related to the effectiveness, including seed type for sprouts, bacterial strain type infecting sprouts, etc. So the treatment type will be the most important predictor that I will focus on. And I am willing to explore if there is any treatment that could reduce bacterial population level on seedling sprouts.

# 3. Methods

Since the dataset is quite well-organized and no any value missing, the cleaning step is relatively easy. I only changed the category names.

My main interest is to test if the chemical treatment is effective or not compared to the control treated groups. Also, anova model will be applied to check which variable will make significant different to the bacterial population. So boxplot, bar chart and line chart will be created according to the analysis direction.

## 3.1 Schematic of workflow

## 3.2 Data acquisition

All of the data in this dataset was collected by me in 2022 after experiment was conducted.

## 3.3 Data import and cleaning

The detailed raw data and processed data are stored in ‘data’ folder. And the code for data cleaning is in the ‘processing-code’ subfolder under ‘R’ folder. Since the name of the categories in variable “Day” is a bit confusing, I changed it to the form of day+time point0,1,3,5,7.

And the reason why I only choose the colony count from selective media is that supportive media may have more colonies that are not specific target strain due to its ingredient feature. So the count from selective media could be more reliable.

## 3.4 Statistical analysis

By fitting the data with ANOVA model, the result will show whether each independent variable has main effect to the dependent variable through the P value. And the summary table is stored in ‘table’ subfolder under ‘result’ folder.

Then for the efficacy of chemical, a LSD test is applied and the significant level will be presented for the 4 treatments.

Bar chart, box plot will show the difference of bacterial populations between 2 types of seeds, 2 types of strains, and among 4 treatments. As well as line graph for the bacteria growth trend during sprouting process.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

My main interest here is the realtionship between treatment and bacterial populations. Based on the ANOVA table, seed type, treatment and day are the variables that have significant effect on the dependent variables; while strain type is not. And there is no significant difference between 2 replications, which means the experiment mayvbe replicable. Furthermore, I ran a LSD test to check the significant level for each 4 treatment and got the result as follows.

Table 1. The result of ANOVA model.

| Variable | Df | Sum Sq | Mean Sq | F value | Pr(F) |
| --- | --- | --- | --- | --- | --- |
| Seed | 1 | 95.8 | 95.81 | 43.287 | 7.53e-10 |
| Strain | 1 | 0.1 | 0.10 | 0.047 | 0.829 |
| Treatment | 3 | 322.3 | 107.43 | 48.538 | < 2e-16 |
| Day | 4 | 113.8 | 28.46 | 12.859 | 5.12e-09 |
| Rep | 1 | 0.4 | 0.42 | 0.190 | 0.664 |
| Residuals | 149 | 329.8 | 2.21 |  |  |

Table 2. Treatment summary table.

| Treatment type | Mean population | Significant level |
| --- | --- | --- |
| Control 2 | 4.79 | a |
| Control 1 | 4.49 | a |
| Chemical treated 2 | 3.23 | b |
| Chemical treated 1 | 1.18 | c |

The figures created are for some basic concepts about the population difference among the categories.

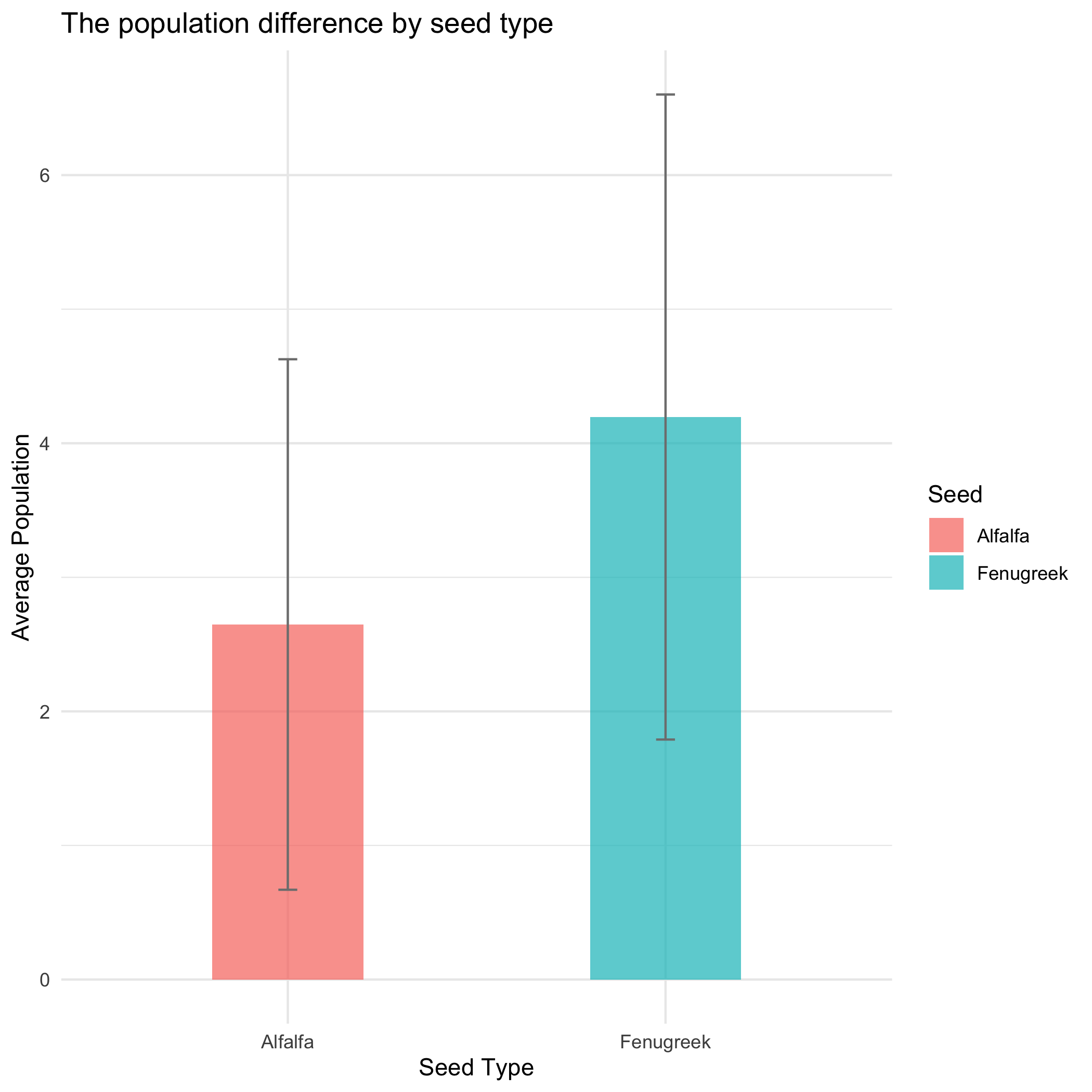


Figure 1. The population difference between alfalfa and fenugreek seeds.

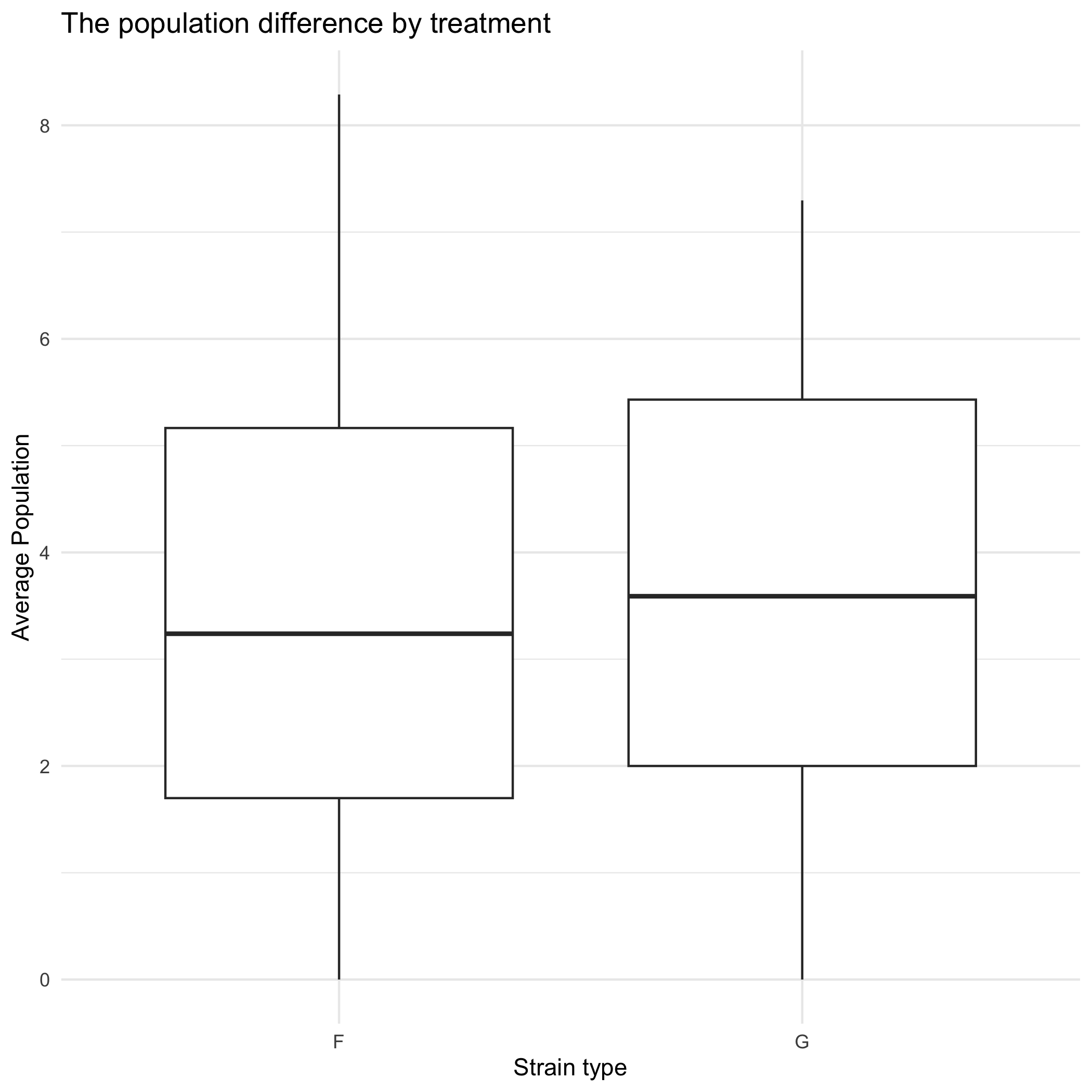


Figure 2. The population difference between F strain and G strain (no significant difference).

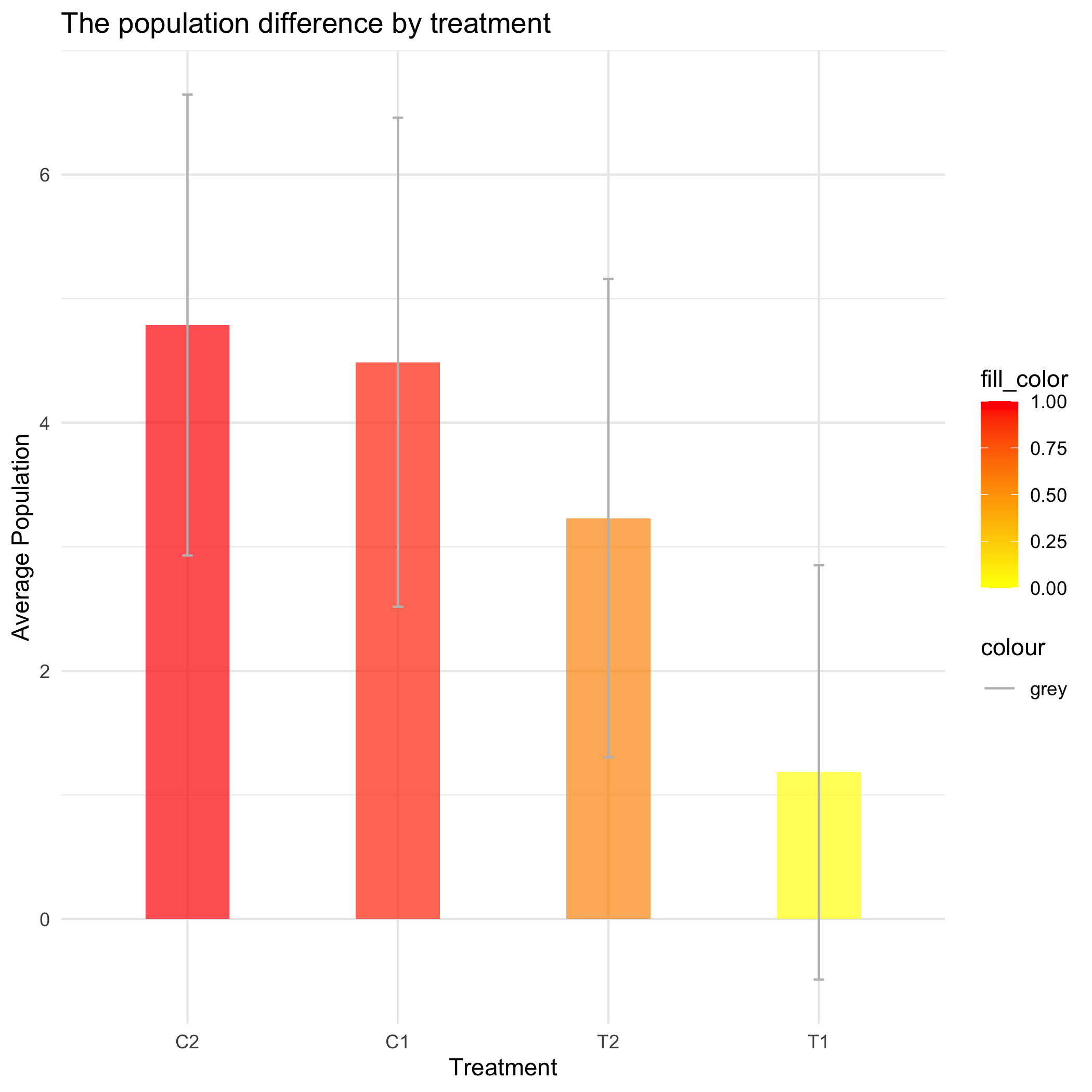


Figure 3. The effect of the chemical treatment application representing by color.

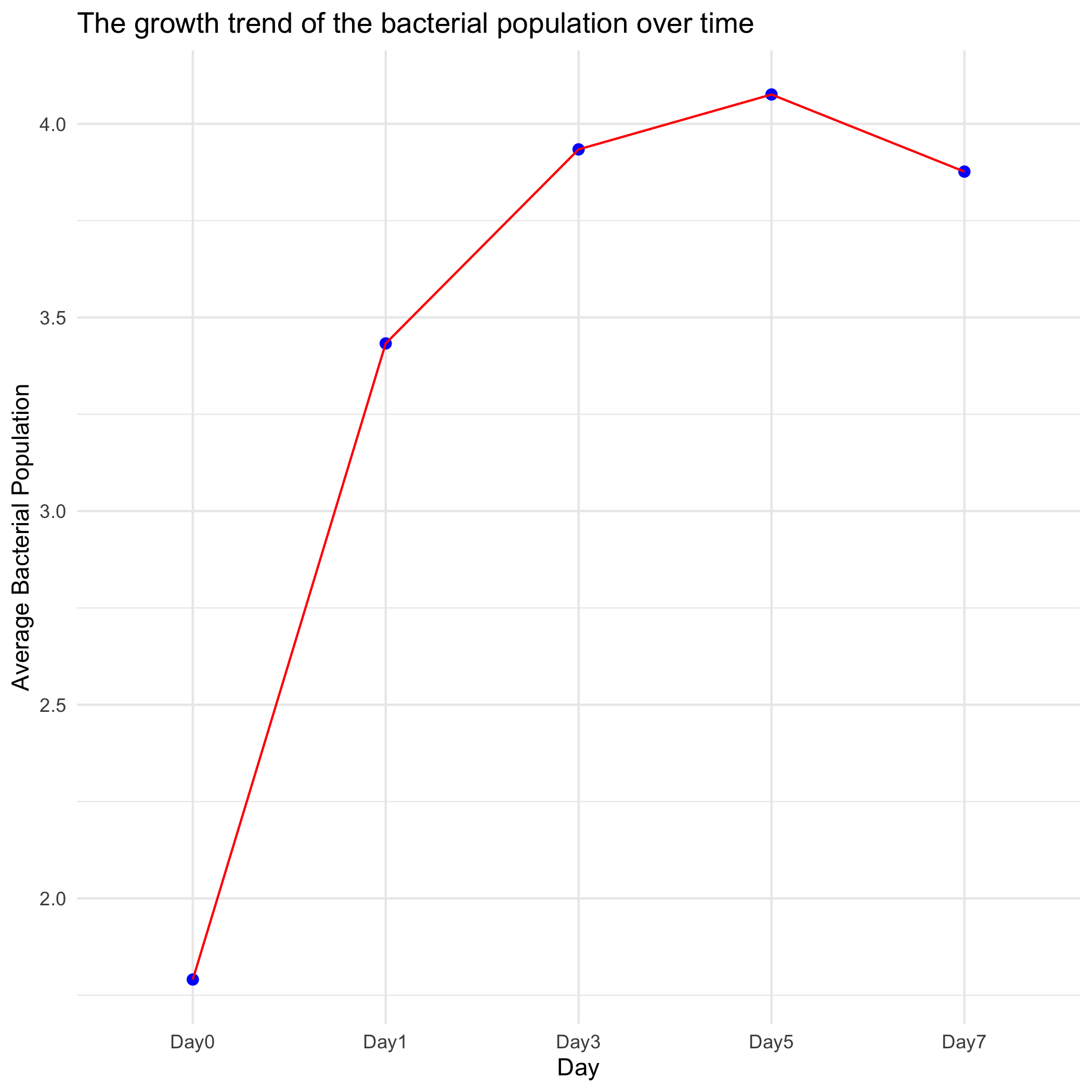


Figure 4. The growth trend of bacteria population during sprouting process.

## 4.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

## 4.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (Leek & Peng, 2015) discusses types of analyses.

These papers (McKay, Ebell, Billings, et al., 2020; McKay, Ebell, Dale, Shen, & Handel, 2020) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 6. References

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*(6228), 1314–1315. <https://doi.org/10.1126/science.aaa6146>

Manosalva, P., Manohar, M., Reuss, S. H. von, Chen, S., Koch, A., Kaplan, F., … Klessig, D. F. (2015). Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. *Nature Communications*, *6*(1), 7795. <https://doi.org/10.1038/ncomms8795>

McKay, B., Ebell, M., Billings, W. Z., Dale, A. P., Shen, Y., & Handel, A. (2020). Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery. *Open Forum Infectious Diseases*, *7*(11), ofaa494. <https://doi.org/10.1093/ofid/ofaa494>

McKay, B., Ebell, M., Dale, A. P., Shen, Y., & Handel, A. (2020). Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients. *Proceedings. Biological Sciences*, *287*(1927), 20200496. <https://doi.org/10.1098/rspb.2020.0496>

Miyahira, R. F., & Antunes, A. E. C. (2021). Bacteriological safety of sprouts: A brief review. *International Journal of Food Microbiology*, *352*, 109266. <https://doi.org/10.1016/j.ijfoodmicro.2021.109266>