Statistics assignment

Sean van der Merwe

February 2023

Instructions

In this experiment, two preservatives (A and B) are being compared. The same contaminated foodstuffs, which differ only in the preservative used, are exposed to 4 carbon substrates. Over time growth occurs and is measured in terms of colour development (CD). The question arises as to which preservative appears to have performed better in this experiment.

1. Find the sheet in the given Excel file that corresponds with your student number or email stub and load the data in a statistical analysis program. [5]
2. Illustrate the full data set any way you prefer that allows you to see and understand the patterns in the data. *Hint: the data is already in the correct form for a ggplot line graph.* [10]
3. Describe the patterns you see. Does one preservative look better to you? How sure are you about your statements? [10]
4. Isolate the values at Hour 48 only, leaving only a single value for each line (8 numbers). Give these numbers in a neat and labelled table. [7]
5. Draw a box plot of these numbers, split by preservative (Treatment A on the left or top and Treatment B on the right or bottom). Which preservative has the lower (better) median? [8]
6. Do a Kruskall-Wallis test for whether there is a statistically significant difference between the preservatives (assuming the substrates are interchangeable) and discuss the results as applicable. Then go on to explain why this was not an appropriate thing to do in this case and/or the various problems with doing the test this way. [20]
7. Draw a heat map of these numbers. *Note that this requires you to put the numbers in a matrix first.* [5]
8. Create a copy of these numbers, but with anything less than 15 replaced with 0, and show the new copy. [5]
9. Calculate the following statistics for each preservative using the adjusted copy: average colour development (mean) and richness (count of how many are above 0). *These calculations must be done with code. While you can and should calculate these by hand to check in this simple case, in full size experiments there are many more numbers and code will be much more efficient.* [10]
10. Looking now at all the plots, tables, and statistics, make a statement as to what you think the results of the experiment are (if anything). [5]
11. Also explain what you personally learned from this assignment (if anything). [5]

Please submit a Word or PDF file with the code, output, and explanations neatly linked and clearly identified. Also attach supporting files as appropriate *(e.g. Rmd or Excel files that you created)*. These must be submitted by emailing to *vandermerwes@ufs.ac.za* at least 24 hours prior to the follow up discussion session. [10]