Major Project

Johann Memmel

150943320

April 6, 2020

Intro

In toxicology, understanding the impacts that a chemical has at a given concentration is critical. To further this goal toxicity tests are carried out in which test organisms are exposed to a known concentration of a chemical in order to see the effects. By comparing the impacts of different concentrations of the chemical we can determine what levels are harmful and what levels don’t have any observable effect. In order to do this NOEC (no observed effect concentration) and LOEC (lowest observed effect concentration) are used. The LOEC is the lowest concentration that results in statistically significant results from the control. The NOEC is the highest concentration that was not significantly different from the control (one concentration lower than the LOEC).

The goal this report will be to determine the NOEC and LOEC of an acute (48 hour) toxicity test of sodium nitrate of *Daphnia magna* in a hardwater media. In order to do this an analysis of variance (ANOVA) will be performed in order to separate the data into groups based on the exposure concentration and to determine if any are significantly different. After, two different tests will be used, the Tukey test and the Dunnett test, in order to determine which concentration is the first to be significantly different from the control. Based on a cursory glance at the data it is clear that at the highest concentrations there was a large impact but at the moderate concentrations it is difficult to say where the differences would begin to be significant. Given that I hypothesize that the NOEC would be at 7g/L and the LOEC would be at 10g/L.

Background

The two different statistical approaches we will use are the Tukey test and the Dunnett test. These two methods are both multiple comparison procedures that are similar to running multiple t-tests. A t-test is a test that is used to compare whether two means are different from each other. By finding the difference between the group means and dividing by the variability, a t-value can be calculated. A significance table is then used with the t-value to determine if the means are statistically significant, meaning that the likelihood of this happening by chance is <5%, giving this test a 5% chance of causing a type 1 error (false positive). In our case which requires multiple comparisons, using multiple t-tests would result in an unacceptable risk of type 1 errors as this risk increases as more tests are done. This is a component of the multiple comparisons problem (which states that as more inferences are made additional false inferences may also occur) called the family-wise error rate (FWER). The FWER is the likelihood of making type one errors when running tests.

In order to get around the multiple comparisons problem and lower the FWER many solutions have been proposed, such as calculating significance thresholds that depend on the number of comparisons being made. This is how the Tukey test corrects for multiple tests, by taking the size of the population the degrees of freedom into consideration when calculating the studentized range (q). This ‘q’ value is then compared the studentized range distribution to determine if the difference is significant. Additionally, the Tukey test also pools variance from the entire data set allowing a more accurate estimate.

The Dunnett test approaches this problem differently starting by reducing the number of comparisons done (k-1 for Dunnett’s vs k(k-1)/2 for Tukey’s where k is the number of trials). This is because the Tukey test compares all possible combinations of trials while the Dunnett test only compares means to the control. This is useful for our purposes as only statistical significance relative to the control is used to find the NOEC and LOEC. The Dunnett test then leverages that all its comparisons are to the control by using the same estimate of error variance. This leads to the Dunnett test having a FWER that cannot be larger than the alpha value of the comparisons. The test also takes into consideration the size of the population and the number of trials similar to how the Tukey test does. Both tests are very similar, functioning as multiple t-tests, but each takes a different approach to reducing the family-wise error rate and handling the multiple comparisons problem.

Additionally, both these tests are easy to run in r, requiring only a few lines each although Dunnett’s test needs an outside package as well (we are using PMCMRplus but there are multiple available). The tests also share the same assumptions; that the data is independent, that observations are normally distributed and share a common variance. Therefor all the code that’s needed to run these tests is; downloading the required package, importing the dataset, generating an ANOVA (Analysis of variance), passing the resulting ANOVA to the Tukey test, running the Dunnett test on the dataset, and confirming that the assumptions of the tests are met. Testing that the assumptions is equally simple except for testing that the observations are independent as this is dependent on experiment design and data collection. Testing that the observations at each trial are normally distributed is done by running the Shapiro-Wilk test on the residual of the ANOVA then confirmed by generating a histogram and Q-Q plot of the residuals as well. Testing that the variance is equal across the groups is done by running the Bartlett test and the Levene test with the Bartlett test being the more sensitive to less normal data.

Methods

For this report the data used is a modified version of results collected by Stephanie Marshall. The results were modified to include additional replicates due to original data containing only two replicates. While this is not legitimate data the methodology and statistical tests run are correct.

See the attached R markdown file for annotated code.

Results

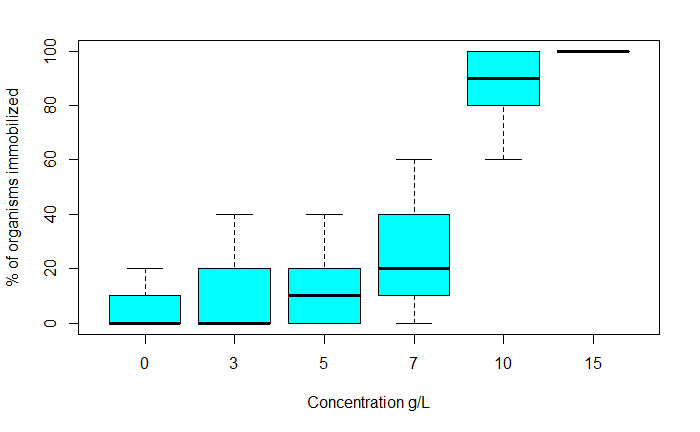


Figure 1. Boxplot showing the percent of organisms immobilized versus the concentration of sodium nitrate. One can clearly see that the percent of organisms immobilized is highest at the higher concentrations. It is unclear from this though what concentration is the first one to be significantly different from the control. Generated by “boxplot(AcuteData[which(AcuteData$Time=='48'),]$Percent.Immobilized~AcuteData[which(AcuteData$Time=='48'),]$Conc, xlab="Concentration g/L", ylab = "% of organisms immobilized", col="cyan")”.

Table 1. The output of the ANOVA generated with ‘g’ being concentration of sodium nitrate. Note that the p-value generated is highly significant indicating that difference between some of the means are significant. Generated by “summary.aov(fit).

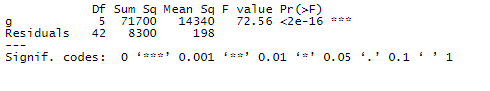


Table 2. The result of the Tukey test. A p adjusted value of <0.05 indicates that the comparison is significant. For our purposes of finding the NOEC and LOEC only the comparisons to the control are of interest. The lowest concentration that led to a significant comparison with the control is the 10g/L concentration. This would make the 10g/L the LOEC and 7g/L the NOEC based on the Tukey results. Generated by “TukeyHSD(fit)”.

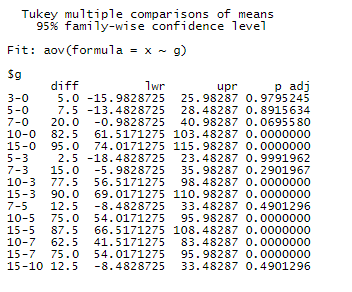
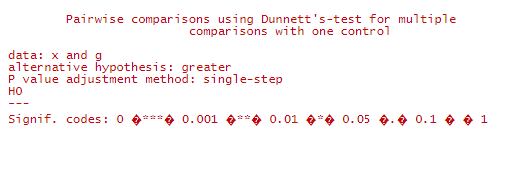


Table 3. The result of the Dunnett test. A p value of <0.05 indicates that the comparison is significant. The lowest concentration that led to a significant comparison was the 7g/L making it LOEC and 5g/L the NOEC for the Dunnett test. Generated by “summary(dunnettTest(x, g, alternative = "greater"))”.



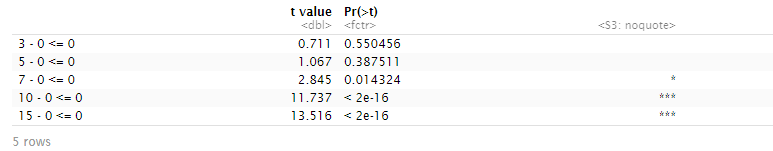
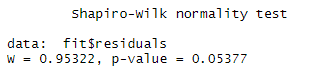


Table 4. Results of the Shapiro-Wilk test on the residuals of the ANOVA. A p value of <0.05 rejects the null hypothesis, that the data is normally distributed. Therefore, in this test as the p-value is 0.05377 the null hypothesis cannot be rejected. Due to the borderline nature of the p-value and the possible impact of sample size additional test will be done. Generated by “shapiro.test(fit$residuals)”.



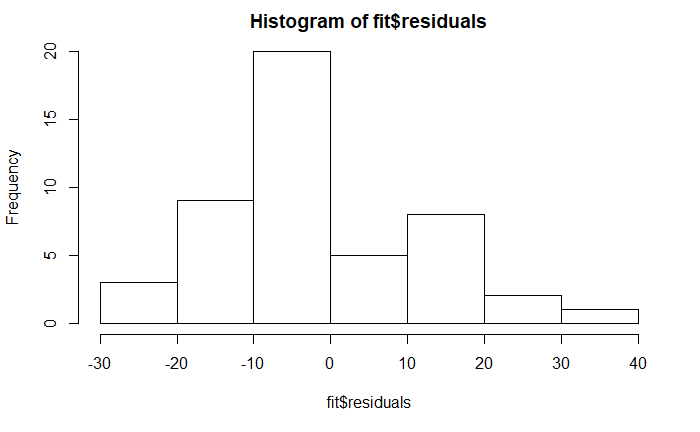


Figure 2. A histogram of the residuals of the ANOVA. While not perfectly normal based on the shape of the chart it makes sense that the Shapiro-Wilk test gave a borderline result. Generated by “hist(fit$residuals)”.

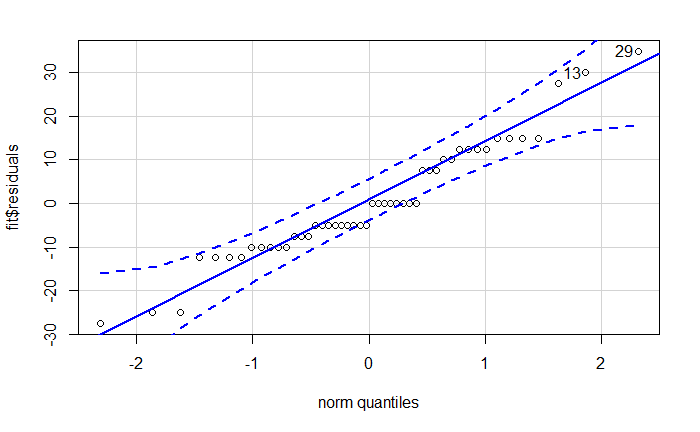


Figure 3. Q-Q plot of the residuals of the ANOVA. Based on this Q-Q plot we observe a relatively linear trend. Due to both the concentration and the response (immobilized %) being discreate variables there are clear “runs” of points in the plot. We also see some outliers tagged.

Generated by “qqPlot(fit$residuals)”.

Table 5. Bartlett test results. A p-value of <0.05 means we cannot reject the null hypothesis stating that the trial variances are all equal. Generated by “bartlett.test(x ~ g)”.

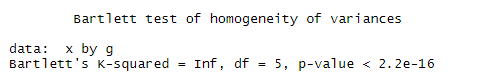
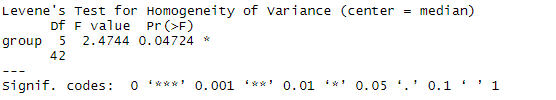


Table 6. Levenes test results. A p-value of <0.05 means we cannot reject the null hypothesis stating that the trial variances are all equal. Generated by “leveneTest(x,g)”.



Discussion

In order to determine the NOEC and LOEC we need to find the lowest concentration that is significantly different from the control. In order to get a rough idea of what the data looks like we generate a boxplot (figure 1.). This is valuable because it can give us an idea of the difference in mean and range of each concentration allowing us to predict that the 10g/L and 15g/L sodium nitrate are likely significant compared to the control while 7g/L is ambiguous about whether it would be. In order to be more confident about which concentrations are significant an ANOVA is run (table 1.). This ANOVA comes back significant with a p-value of <2e-16 significantly lower than the <0.05 alpha value needed to reject the hypothesis that the means are all the same. Knowing that at least one comparison is significant now we move on to figuring out which comparisons. We do this by running the Tukey test and the Dunnett test. This is where the two methodologies differ.

The Tukey test results (table 2.) indicate that multiple comparisons are significant as multiple p-adjusted values are <0.05. For our purposes of finding the NOEC and LOEC only the comparisons to the control are of interest. The lowest concentration that led to a significant comparison with the control is the 10g/L concentration. The 7g/L comparison with the control resulted in a p-adjusted value of 0.069 falling just outside the significant range. This would make the 10g/L the LOEC and 7g/L the NOEC.

The Dunnett test results (table 3.) shows that 10, 15 and 7g/L all have p-values of <0.05 when compared to the control, indicating that the comparison is significant. The lowest concentration that led to a significant comparison was the 7g/L making it LOEC and 5g/L the NOEC.

The assumptions of these two methods are the same and therefore can be tested with the same methods. There is no way to ensure the independence of data as requires investigation into the experiment design and that is unfeasible for this experiment. To test that the distribution at each concentration is normally distributed a Shapiro-Wilk test was used (table 4.). This test reported a p-value of 0.05377 just slightly above the critical value that would allow us to meet this assumption. To confirm this result a histogram (figure 2.) and a Q-Q plot (figure 3.) of the residuals will be used. The histogram allows up to observe the shape and determine if it looks ‘normal’, this case appearing reasonably normal. The Q-Q plot also looks reasonably good admittedly with odd ‘runs’ due the concentration and the response (immobilized %) being discreate variables. Based on these results I would accept that the data is reasonably normal, meeting the second assumption. In order to test the there are equal variances across the different concentrations the Bartlett test (table 5.) and Levene’s test (table 6.) will be used. Both of these tests return p-values that are less than <0.05 meaning we can reject the null hypothesis that the variances in each group are the same. Therefore, our data fails the third assumption. While this is problematic. Due to the nature of the results where not less than 0% of *Daphnia* can be immobilized and where not more than 100% of *Daphnia* its likely that the variances would not be equal across the concentrations.

With the Tukey and Dunnett tests in disagreement we are left to decide which one to believe. The results of the Tukey test say that 7g/L is not significantly different from the control with a close p-value of 0.069. While the Dunnett test says that this comparison is significant with a p-value of 0.014, safely below the critical threshold of 0.05. Based on the Tukey tests more general nature I would lean towards trusting the Dunnett results more. The Dunnett test is specialty designed for comparing trials to the control and therefore is more tailored to our application. Therefore, we conclude that the LOEC is 7g/L and the NOEC is 5g/L sodium nitrate.

Both the Tukey test and Dunnett test have useful applications. For this specific application where only comparisons against the control are needed the Dunnett test is the ideal test to use. It enabled us to determine the LOEC and NOEC of sodium nitrate on *Daphnia magna*. While the assumptions of these tests proved to be problematic with our dataset they are straightforward to test.

Citations

NIST. 2013. NIST/SEMATECH e-Handbook of Statistical Methods. National Institute of Standards and Technology. From <https://www.itl.nist.gov/div898/handbook/prc/section4/prc471.htm>

Dunnett C. W. 1955. "A multiple comparison procedure for comparing several treatments with a control". Journal of the American Statistical Association. 50: 1096–1121.

OECD. 2014. “Current Approaches in the Statistical Analysis of Ecotoxicity Data”. Organisation for Economic Co-operation and Development

Howell C. 2010 "Statistical Methods for Psychology 7th edition”. From <https://labs.la.utexas.edu/gilden/files/2016/05/Statistics-Text.pdf>

Laerd Statistics. “Hypothesis Testing”. From <https://statistics.laerd.com/statistical-guides/hypothesis-testing.php>

Tukey J. 1949. “Comparing Individual Means in the Analysis of Variance”. Biometrics. 5:2, 99-114

RPubs. “Post-Hoc Analysis with Tukey’s Test”. Rpubs by RSudio. From <https://rpubs.com/aaronsc32/post-hoc-analysis-tukey>

Pohlert T. 2020. “PMCMRplus documentation”. Rdrr.io from <https://rdrr.io/cran/PMCMRplus/man/dunnettTest.html>

Rumsey, Deborah 2009. “Statistics II for Dummies”. Wiley. pp. 186 from <https://archive.org/details/isbn_9780470466469/page/186>

STHDA. “One-Way ANOVA Test in R” from <http://www.sthda.com/english/wiki/one-way-anova-test-in-r>

GraphPad. “Tukey and Dunnett methods”. From <https://www.graphpad.com/guides/prism/7/statistics/stat_the_methods_of_tukey_and_dunne.htm>

Daniel J. 2016 “Ecotoxicology – NOEC and LOEC” from <https://www.quantics.co.uk/blog/ecotoxicology-noec-and-loec/>