Major Project

Johann Memmel

150943320

April 6, 2020

Intro

In toxicology, understanding the impacts that a chemical has at different concentrations 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, we can determine what levels are harmful and what levels do not have any observable effect. In order to explain this the terms 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 is not significantly different from the control (one concentration lower than the LOEC).

The goal this report is to determine the NOEC and LOEC of an acute (48 hour) toxicity test of sodium nitrate on *Daphnia magna* neonates in a hardwater media. To do this, an analysis of variance (ANOVA) will be performed. This will separate the data into groups based on the exposure concentration and then determine if there are any significant differences between them. Then 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, at the highest concentrations there was a significant difference but at lower concentrations it is unclear where 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 I will use are the Tukey test and the Dunnett test. These two methods are both multiple comparison procedures 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 this case which requires multiple comparisons, using multiple t-tests would result in an unacceptable risk of type 1 errors. As the risk of type 1 errors 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 comparisons, by taking the size of the population and the degrees of freedom into consideration when calculating the studentized range (q). This ‘q’ value is then compared to 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. Dunnett’s test makes k-1 comparisons while the Tukey test does k(k-1)/2 comparisons 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 (I am using PMCMRplus but multiple packages are available). The tests also share the same assumptions; that the data is independent, the observations are normally distributed and share a common variance. Therefore all 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 the assumptions is equally simple excluding the independence of observations. Determining if the observations are independent is difficult as it depends on experiment design and data collection. Testing that concentration is normally distributed is done by running the Shapiro-Wilk test on the residual of the ANOVA. Then by confirming with a histogram and Q-Q plot on 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

The data used for this report 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 or see appendix 1.

Results

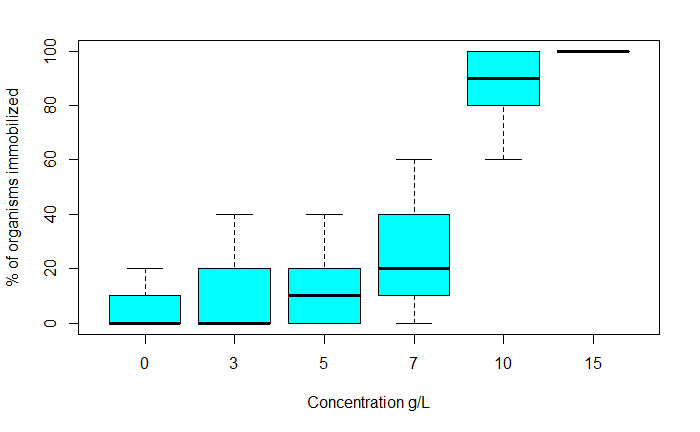


Figure 1. Boxplot showing the percent of organisms immobilized versus the concentration of sodium nitrate. One can clearly see that the percent of organisms immobilized is highest at the higher concentrations. It is unclear from this though what concentration is the first one to be significantly different 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).

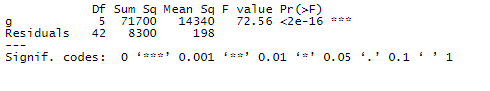


Table 2. The result of the Tukey test. A p adjusted value of <0.05 indicates that the comparison is significant. For our purposes of finding the NOEC and LOEC only the comparisons to the control are of interest. The lowest concentration that led to a significant comparison with the control is the 10g/L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

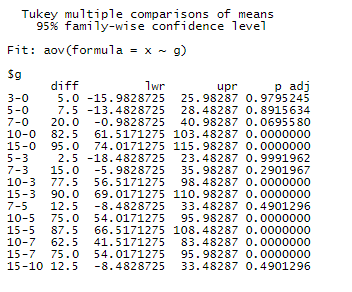
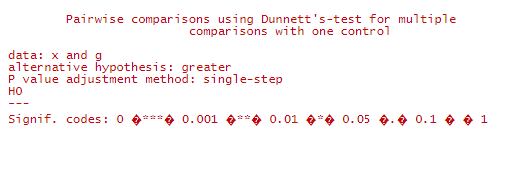


Table 3. The result of the Dunnett test. A p value of <0.05 indicates that the comparison is significant. The lowest concentration that led to a significant comparison was the 7g/L making it LOEC and 5g/L the NOEC for the Dunnett test. Generated by “summary(dunnettTest(x, g, alternative = "greater"))”.



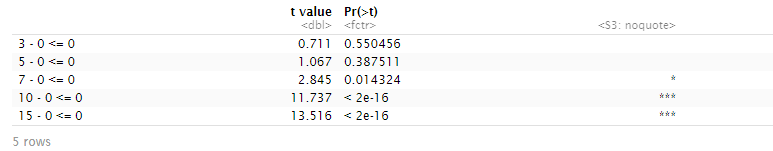
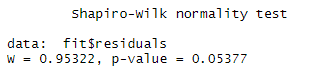


Table 4. Results of the Shapiro-Wilk test on the residuals of the ANOVA. A p value of <0.05 rejects the null hypothesis, that the data is normally distributed. Therefore, in this test as the p-value is 0.05377 the null hypothesis cannot be rejected. Due to the borderline nature of the p-value and the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



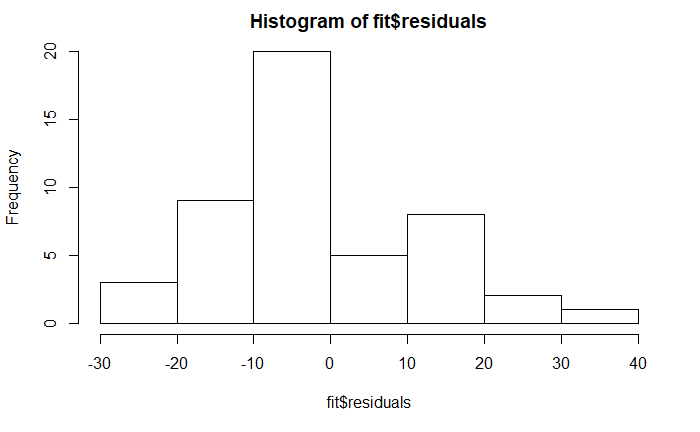


Figure 2. A histogram of the residuals of the ANOVA. While not perfectly normal based on the shape of the chart it makes sense that the Shapiro-Wilk test gave a borderline result. Generated by “hist(fit$residuals)”.

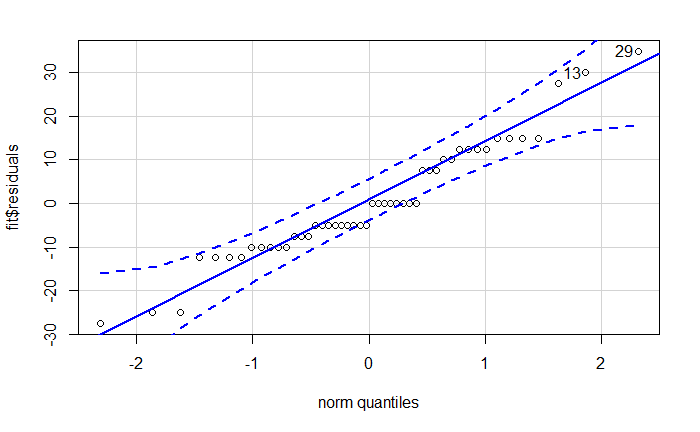


Figure 3. Q-Q plot of the residuals of the ANOVA. Based on this Q-Q plot we observe a relatively linear trend. Due to both the concentration and the response (immobilized %) being discreate variables there are clear “runs” of points in the plot. We also see some outliers tagged.

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)”.

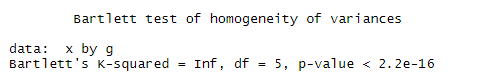
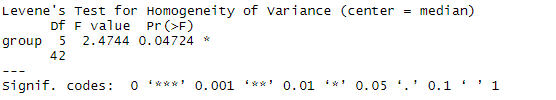


Table 6. Levenes test results. A p-value of <0.05 means we cannot reject the null hypothesis stating that the trial variances are all equal. Generated by “leveneTest(x,g)”.



Discussion

In order to determine the NOEC and LOEC the lowest concentration that was significantly different from the control was needed. To get a rough idea of what the data looked like, a boxplot was generated (Figure 1.). This is valuable because it gives an idea of the mean and range of each concentration. Based on this is appears that the 10g/L and 15g/L sodium nitrate concentrations are likely significant. While the 7g/L concentration was ambiguous about whether is significant. In order to be more confident about which concentrations are significant an ANOVA was run (Table 1.). This ANOVA came back significant with a p-value of <2e-16, which is 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 I can continue on to determine which specific comparisons are significant. This can be done by running the Tukey test or the Dunnett test. In order to compare the results of these two tests both will be run.

The Tukey test results (Table 2.) indicated that multiple comparisons are significant as multiple p-adjusted values are <0.05. For the 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 are the 10g/L concentration. The 7g/L comparison with the control resulted in the 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.) showed 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 it requires investigation into the experiment design. 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 one to observe the shape and determine if it looks ‘normal’. In this case, the histogram appears 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 that there are equal variances across the different concentrations, the Bartlett test (Table 5.) and Levene’s test (Table 6.) were used. Both of these tests returned p-values that are less than <0.05 meaning the null hypothesis that the variances in each group are the same can be rejected. Therefore, the data used fails the third assumption. This was an unavoidable problem due to the nature of the results. As the response variable cannot be less then 0% of *Daphnia* immobilized and not higher than 100% it was likely that the variances be different at the extremes.

With the Tukey and Dunnett tests in disagreement one is 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 this application. Therefore, I 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. While both tests could determine the LOEC and NOEC of sodium nitrate on *Daphnia magna* the Dunnett test results are more trustworthy. While the assumptions of these tests proved to be problematic with this dataset, they are straightforward to test and could be avoided by more replicates.

Citations

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

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

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

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>

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>

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

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

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

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>

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

Appendix 1. R Code

---

title: "Final Project"

author: "Johann Memmel"

date: "27/03/2020"

output: html\_document

---

```{r}

if (!require("PMCMRplus")) install.packages("PMCMRplus")

library(PMCMRplus)

if (!require("car")) install.packages("car")

library(car)

#downloads/installs packages and loads them

```

```{r}

AcuteData = read.csv("48hr acute Sodium Nitrate test mod.csv", fileEncoding="UTF-8-BOM")

#Loads the csv file

```

```{r}

boxplot(AcuteData[which(AcuteData$Time=='48'),]$Percent.Immobilized~AcuteData[which(AcuteData$Time=='48'),]$Conc, xlab="Concentration g/L", ylab = "% of organisms immobilized", col="cyan")

#Generates a boxplot of the concentration vs the percent immobilized

#figure 1

```

```{r}

x <- AcuteData[which(AcuteData$Time=='48'),]$Percent.Immobilized

g <- as.factor(AcuteData[which(AcuteData$Time=='48'),]$Conc)

#sets the variables x and g to the percent immobilized and the concentration at time = 48. This is because we are only intrested in the final results and not the results at time = 0 or time = 24

fit <- aov(x ~ g)

#fits the variables into an ANOVA

```

```{r}

summary.aov(fit)

#tests the ANOVA to show if all the means are equal. A significant results indicates that not all means are equal.

#table 1

```

```{r}

TukeyHSD(fit)

#Runs the Tukey test on the ANOVA. a p adjusted value of <0.05 indicates a significant diference

#table 2

```

```{r}

summary(dunnettTest(x, g, alternative = "greater"))

#Runs the Dunnett test. A p value of <0.05 indicates a significant diference.

#Table 3

```

```{r}

shapiro.test(fit$residuals)

#runs shapiro wilk test to determine if the residuals are normally distributed6

#Table 4

hist(fit$residuals)

#Generates a histogram to confirm shapiro wilks test result

#figure 2

qqPlot(fit$residuals)

#Generates a Q-Q plot to confirm shapiro wilks test result

#figure 3

```

```{r}

bartlett.test(x ~ g)

#table 5

leveneTest(x,g)

#table 6

#these two tests confirm that the amount of variance in each comparison group is equal

```