ANOVA Tutorial

Simon Hans Edasi

2023-05-31

Section 5.6

The Analysis of Variance (ANOVA)

In the classic one-way ANOVA, the explanatory variable is a categorical variable which has two or more levels, e.g., recall the Grazing variable in the compensation data set. The Grazing variable had two levels, Grazed and Ungrazed.

We will follow along with BC&P. The dataset we are going to use centers on water fleas, Daphnia spp., and their parasites.

The question we are asking focuses on water flea growth rates and has two parts.

1. Generally, do parasites alter the growth rates of Daphnia?

2. The data come from a well-replicated and designed experiment, we can also ask whether ach of three parasites reduces growth, compared with a control, no parasite treatment.

```
# Clear the memory
rm(list = ls())


# Load libraries
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
```

3. Get the Daphniagrowth.csv dataset.The data is structured as a data frame.In this dataset, growth.rate is the dependent (or response) variable and parasite is the independent (predictor) variable.

To make the plotting of the results from the analysis of this data set easier, it has been modified from what BC&P created. An additional column has been added that provides an abbreviation for the genus and species of each parasite used in the study. Running View() will display the new structure of the data set. The new column is named "parasite". the full taxonomic names appear in the column named "parasite.name".

```
daphnia = read.csv('Daphniagrowth.csv')
```

```
View(daphnia)
```

Look at it using two functions: View(), and head()

```
##   parasite.name parasite rep growth.rate
## 1       control  control   1    1.074709
## 2       control  control   2    1.265902
## 3       control  control   3    1.315156
## 4       control  control   4    1.075752
## 5       control  control   5    1.196762
## 6       control  control   6    1.383795
```

Since this course is focused on use of statistics in the biological sciences, it is reasonable to ask what kind of parasites are these? You should find out what kind of organism Daphnia is.

An internet search provides the following information:

1. Pasteuria ramosa is a gram-positive, endospore-forming bacterium in the Bacillus/Clostridia clade within Firmicutes. It is an obligate pathogen of cladoceran crustaceans from the genus Daphnia. An established and widely used coevolutionary model of host-pathogen interactions exists with P. ramosa and D. magna.

2. Pansporella perplexa is an amoeboid parasite. Look it up in relation to D. magna.

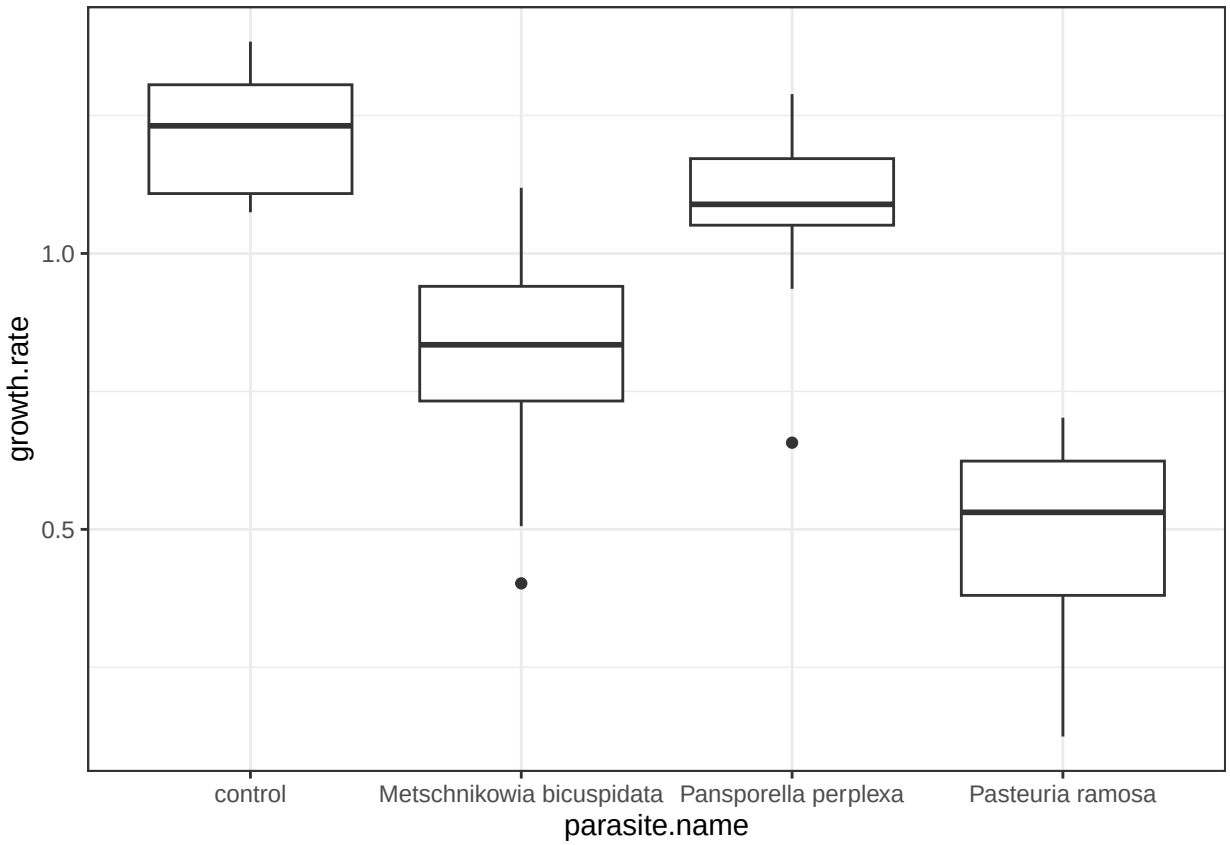
3. etschnikowia are single-celled fungal parasites of freshwater animals. It typically parasitises crustaceans, including Daphnia, a genus of zooplankton. M. bicuspidata has also been found to infect prawns and salmon.

So, we have three different types of parasites; a bacterium, an amoeba, and a fungus.

We can see that the data frame has four variables, two of which we want to use for figure making, growth.rate and parasite. The other two, rep and parasite.name, indicate the replication in each treatment level and the full taxonomic name of each parasite, respectively.

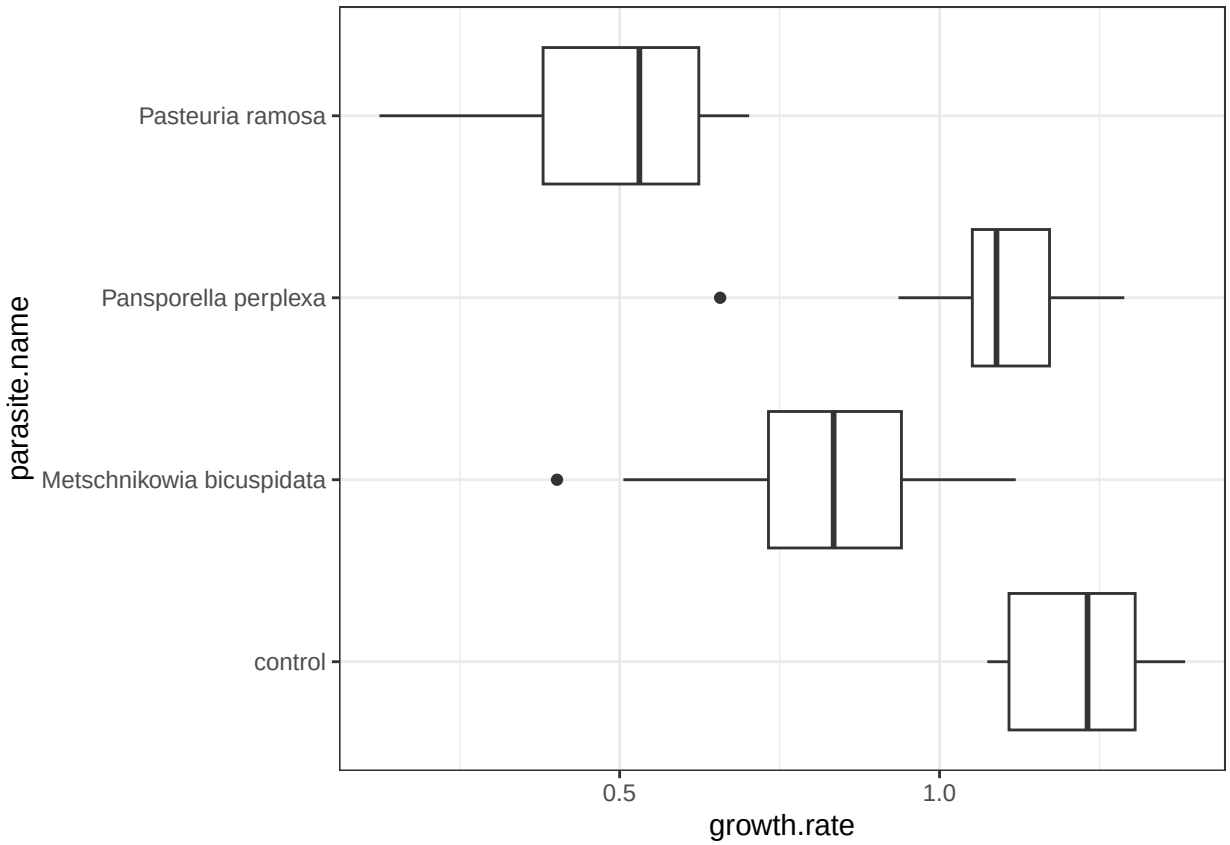
In Chapter 4, where we introduced the box-and-whisker plot as a quick and effective tool for viewing variation in a response variable as a function of a grouping, categorical variable? That's probably the start we want to make.

4. Create a box plot of the daphnia data.



5. Use one method to fix the graph for better understanding.

```
ggplot(daphnia, aes(x = parasite.name, y = growth.rate)) +
geom_boxplot() +
theme_bw() +
coord_flip()
```



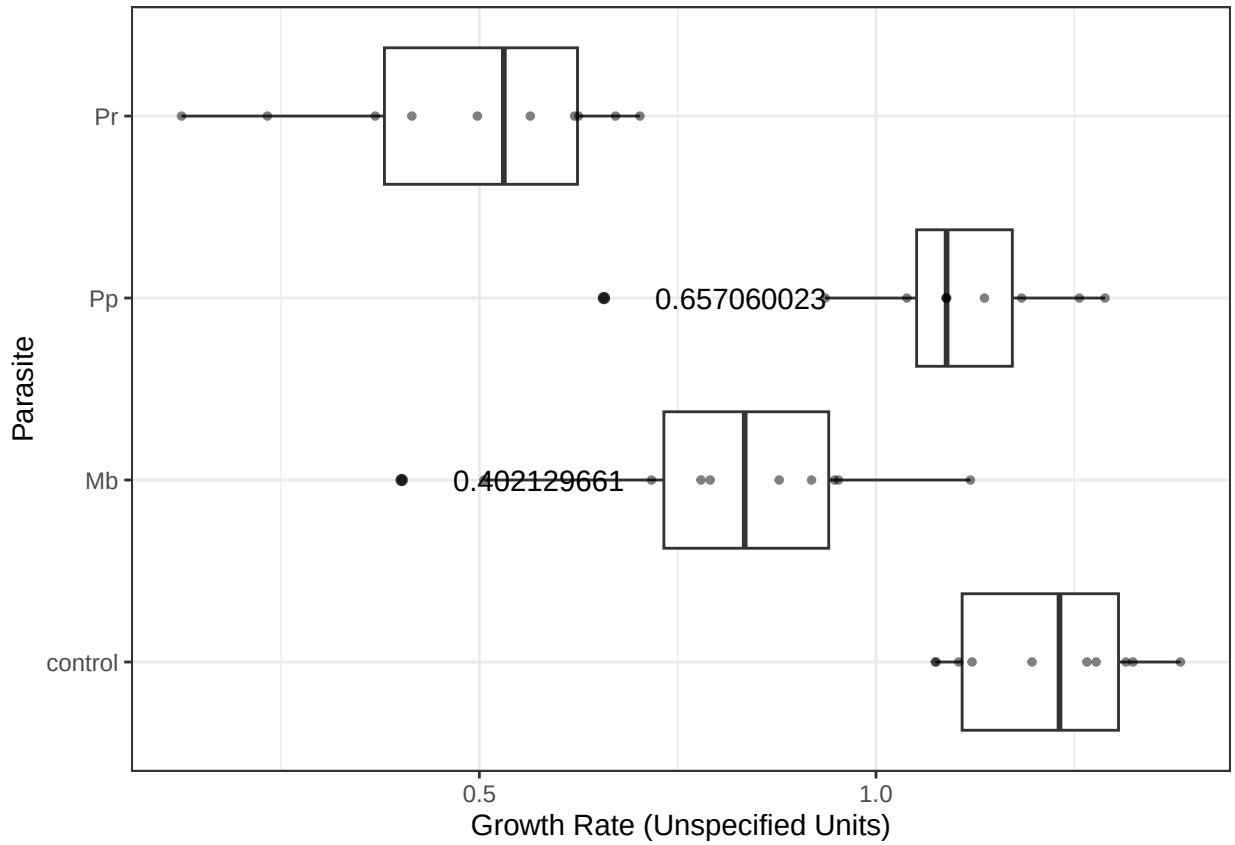
Let's search for outliers and then plot the box plots. The script relies on an often used mathematical definition of what might be an outlier. Ultimately, the determination of what constitutes an outlier rests with the investigator and their knowledge of their subject. A mathematical method unthinkingly employed is insufficient to determine an outlier. The method is used here as a scanning tool to find potential outliers.

This script introduces a user created R function named "is_outlier". Such functions are highly useful for creating code that can be used many times within a larger R script. Doing this saves rewriting a similar piece of code many times within the larger script. Learning how to create R functions is beyond what we will learn of R in this course, but perhaps you may now be able to dissect the script to see how it works.

```
is_outlier = function(x) {
  return(x < quantile(x,0.25) - 1.5 * IQR(x) |
           x > quantile(x, 0.75) + 1.5 * IQR(x))
}

daphnia %>%
  group_by(parasite) %>%
  mutate(outlier = ifelse(is_outlier(growth.rate), growth.rate, as.numeric(NA))) %>%
  ggplot(., aes(x = factor(parasite), y = growth.rate)) +
  geom_boxplot() +
  geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.3)+
  geom_point(size = 1, colour = 'black', alpha = 0.5) +
  xlab("Parasite") +
```

```
ylab("Growth Rate (Unspecified Units)") +
theme_bw() +
coord_flip()
```



Observations from BC&P

1. We can note that there is substantial variation in the daphnia growth rates among the four treatments.

2. We can see that the control treatment produces the highest growth rate, about 1.2 mm/day.

3. We can see that P. perplexa is closest and perhaps lower, M. bicuspidata is next lowest, and P. ramosa is definitely lower.

4. We can see that there is likely to be a parasite treatment effect overall (question 1), and an ordering in the growth rates, with
parasites generally driving down the growth rate, and P. ramosa $<$ M. bicuspidata $<$ P. perplexa (question 2).

5. We can go even further, estimating the average growth rate for each treatment (looks like about 1.2 for the control treatment), and the difference of the parasite treatments from the control treatment (i.e.,the effects of each parasite on the growth rate of Daphnia spp.).

6. Even though BC&P do not do it, we can check some of these comments by calculating a five number summary for the control and each parasite by using the core R function "by()" as follows (first, we will once again view the entire dataset:

```
# View(daphnia)
?by()
```

```
by(daphnia, daphnia$parasite, summary)
```

```
## daphnia$parasite: control
##  parasite.name        parasite             rep         growth.rate
##  Length:10          Length:10          Min.   : 1.00   Min.   :1.075
##  Class :character   Class :character   1st Qu.: 3.25   1st Qu.:1.109
##  Mode  :character   Mode  :character   Median : 5.50   Median :1.231
##                                        Mean   : 5.50   Mean   :1.214
##                                        3rd Qu.: 7.75   3rd Qu.:1.306
##                                        Max.   :10.00   Max.   :1.384
## ------------------------------------------------------------
## daphnia$parasite: Mb
##  parasite.name        parasite             rep         growth.rate
##  Length:10          Length:10          Min.   : 1.00   Min.   :0.4021
##  Class :character   Class :character   1st Qu.: 3.25   1st Qu.:0.7326
##  Mode  :character   Mode  :character   Median : 5.50   Median :0.8345
##                                        Mean   : 5.50   Mean   :0.8012
##                                        3rd Qu.: 7.75   3rd Qu.:0.9404
##                                        Max.   :10.00   Max.   :1.1191
## ------------------------------------------------------------
## daphnia$parasite: Pp
##  parasite.name        parasite             rep         growth.rate
##  Length:10          Length:10          Min.   : 1.00   Min.   :0.6571
##  Class :character   Class :character   1st Qu.: 3.25   1st Qu.:1.0512
##  Mode  :character   Mode  :character   Median : 5.50   Median :1.0890
##                                        Mean   : 5.50   Mean   :1.0764
##                                        3rd Qu.: 7.75   3rd Qu.:1.1719
##                                        Max.   :10.00   Max.   :1.2888
## ------------------------------------------------------------
## daphnia$parasite: Pr
##  parasite.name        parasite             rep         growth.rate
##  Length:10          Length:10          Min.   : 1.00   Min.   :0.1246
##  Class :character   Class :character   1st Qu.: 3.25   1st Qu.:0.3804
##  Mode  :character   Mode  :character   Median : 5.50   Median :0.5309
##                                        Mean   : 5.50   Mean   :0.4822
##                                        3rd Qu.: 7.75   3rd Qu.:0.6237
##                                        Max.   :10.00   Max.   :0.7024
```

We are going to follow BC&P and use the R function "anova()" in the sections on the analysis of variance (ANOVA) and simple linear regression. The core R contains a resident statistical package, "stats". This package contains two analysis of variance functions "aov()" and "anova()". The more general and expandable of these is "anova()" and BC&P introduce it because they make extensive use of it in:

```
Chapter 6: Advancing Your Statistics in R and
Chapter 7: Getting Started with Generalized Linear Models.
```

As this is an introductory course in statistics, we will not study these methods even though they are useful and frequently used in scientific investigations. Please feel free to read these chapters and create the scripts the authors prompt you to write. Our text and several other introductory statistical texts cover two-way ANOVA and reading that in conjunction with BC&P will be helpful in studying these techniques. Both of these topics are covered in depth in Q SCI 482 and 483.

We will follow BC&P and use "anova()" until the end of our analysis where we will look at using Tukey's HSD (Honestly Significant Difference test) to examine which pair(s) of population means are different after completing an ANOVA and finding that we reject the null hypothesis. Here, using "aov()" along with a simple default "plot()" function will easily get us the results we what. Both"anova()" and "aov()" offer the ability to perform Tukey's test, but sometimes the less expansive function is just easier to use to get the job done. That is a good lesson to remember.

7. Compare information on "anova()" and "aov()" by asking for help.

```
# ?anova()
?aov()
```

To perform a one-way ANOVA, we use "anova()" along with the function "lm()" which constructs a linear model we use as an argument in "anova()". Together with "summary()", these other two functions will provide the values we will need to perofm the ANVOA on the Daphnia data.

The ANVOA tests the null hypothesis that all of the groups come from populations that have the same mean. In this case the four groups are the control and the three parasites that infect the Daphnia. In general, we refer to these groups as "treatments". In our experiment, we would call these groups the "parasite treatment".

The test statistic has an F distribution. Its value quantifies the ratio of the between-group variance (between the control and three parasite groups) to the within-group variance (the variance within the control data and within the data for each parasite group).

We will now perform the ANOVA. First, we estimate a linear model from our data that proposes a relationship between the growth rate of Daphnia and their infection with parasites. Then, check to see if the assumptions of the ANOVA are at least approximately met.

8. Fit a linear model for the dependence of growth.rate on parasite.

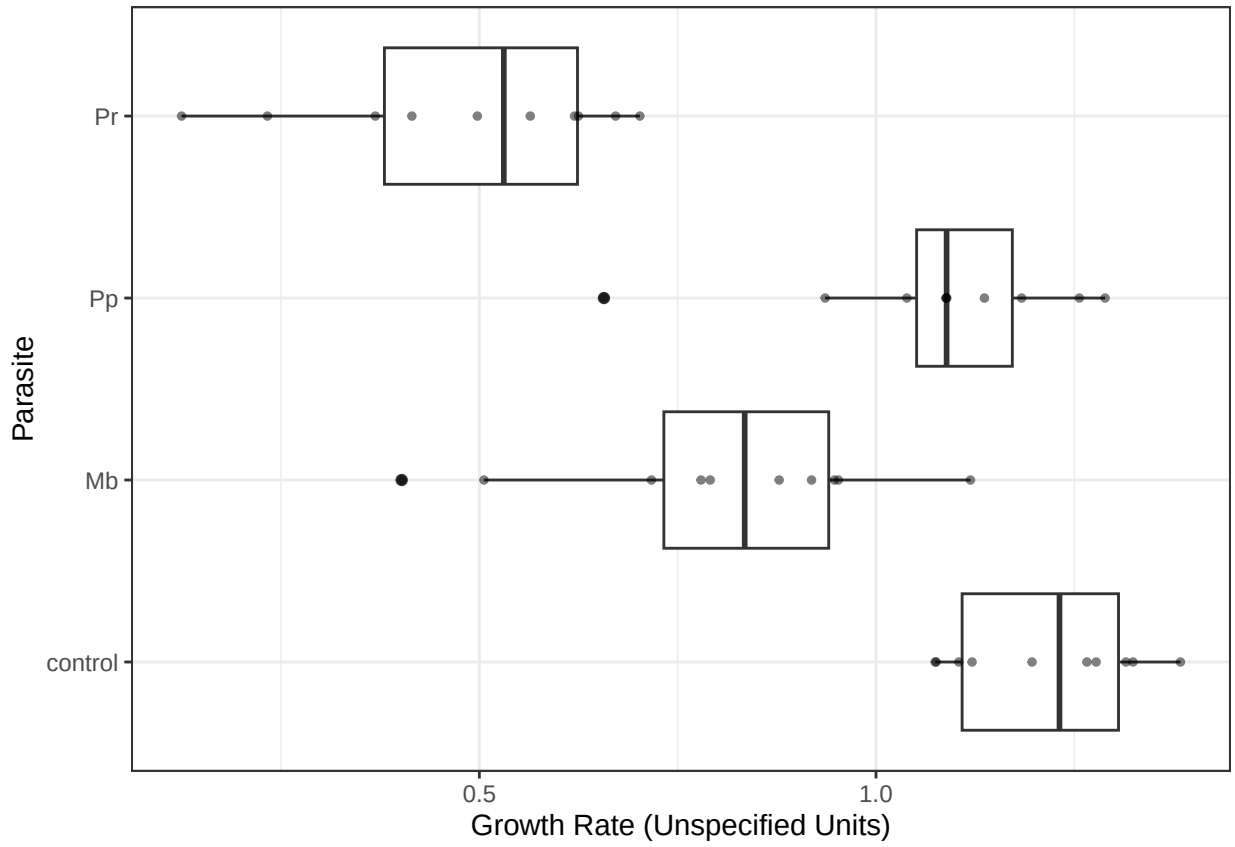
```
model_grow <- lm(growth.rate ~ parasite, data = daphnia)
```

Let's search for outliers and then plot the box plots. Notice that a user created R function has been used in this script. I fwe wrote the script in a contiguous manner, the function could have been created early on in the larger script and called (used) here and anywhere else we might need it.

We will also search for outliers by visually examining the veracity of the assumptions for this experiment using the box plots we did and the autoplot() function in the package "ggfortify".

Be sure to spend some time dissecting what the following script (code) segment is doing.
Do not forget that dissecting code that is already written, and that works, is a good way to learn coding yourself in any software package.

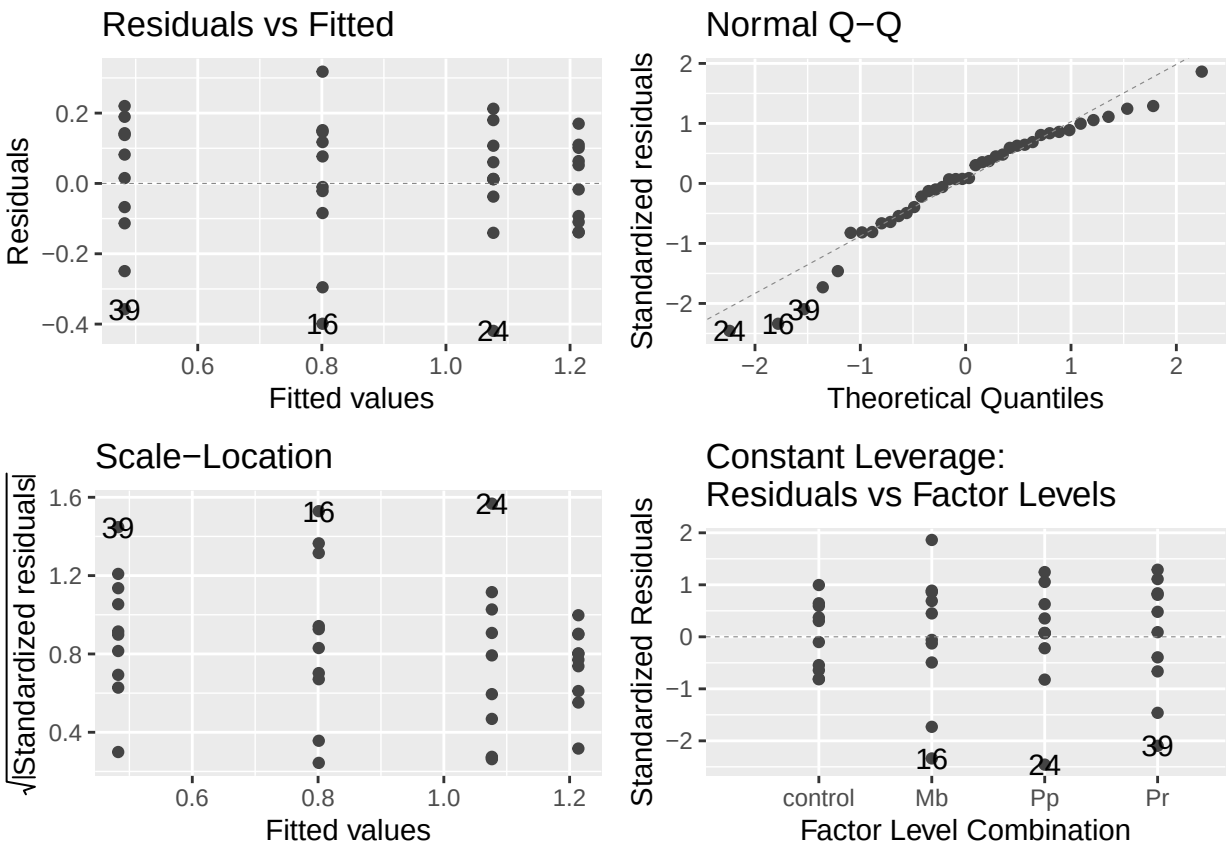
```
is_outlier = function(x) {
  return(x < quantile(x,0.25) - 1.5 * IQR(x) |
           x > quantile(x, 0.73) + 1.5 * IQR(x))
}
daphnia %>%
  group_by(growth.rate) %>%
  mutate(outlier = ifelse(is_outlier(growth.rate), growth.rate, as.numeric(NA))) %>%
  ggplot(., aes(x = factor(parasite), y = growth.rate)) +
  geom_boxplot() +
  geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.3)+
  geom_point(size = 1, colour = 'black', alpha = 0.5) +
  xlab("Parasite") +
  ylab("Growth Rate (Unspecified Units)") +
  theme_bw() +
  coord_flip()
```



9. Plot four diagnostic plots that examine the residuals.

```
# install.packages('ggfortify')
library(ggfortify)
autoplot(model_grow, smooth.colour = NA)
```

```
## Warning: Removed 40 rows containing missing values (`geom_line()`).
## Removed 40 rows containing missing values (`geom_line()`).
```



Let's Examine the Residual Plots

Now examine these plots. All of these plots are based on analyzing residuals-errors.

Interpretation

The interpretation of these plots is (from BC&P):

1. Top left.

This panel is about the 'systematic part' of the model; it tells us whether a line is appropriate to fit to the data. If things have gone wrong, hump-shapes or valleys will be apparent. These would mean that the structure of your model was wrong. For example, fitting a straight line didn't work.

Hence, this plot assesses possible departures from the linear model we fitted and/or consistently changing variance. The ideal pattern would be a horizontal band of residuals scattered on either side of a horizontal line at zero. Watch out for:

a) Curved patterns or other obvious trends in the plot;

b) Fanning patterns showing the variability consistently increasing or decreasing; and

c) Outliers that are unusually far above or below the zero line.

Many people suggest you look at this to evaluate the assumption of equal variance-homo- vs heteroskedasticity (unequal variances). But there is a better plot (bottom left) for this.

2. Top right.

This Q-Q Plot evaluates the assumption of normality of the residuals. The dots are the residuals, and the dashed line the expectation under the normal distribution. This is a much better tool than making a histogram of the residuals, especially with small sample sizes like less than 100.

3. Bottom left.

This evaluates the assumption of equal variance. The y-axis is a standardized (all positive) indicator of the variation. Linear models assume that the variance is constant over all predicted values of the response variable. There should be no pattern. But there might be one if, for example, the variance increases with the mean, as it might with count data (see Chapter 7).

4. Bottom right.

This evaluates leverage, a tool not only to detect influential data points, ones that move the gradient more than might be expected, but also to detect outliers.

As BC&P point out, these figures suggest everything is probably fine. That is, the assumptions of ANOVA are met well enough for use to go on with the analysis. Never forget that it is up to you as the analyst to to justify why the assumptions are met. Others may wish to disagree, so have your agruments we considered.

Using function"anova()"

This function, applied to our simple linear model, will provide an answer
to our first question posed above:

Is there any effect of our treatments (parasite infections and no such infection)? In other words, do parasites infections affect the growth rate of Daphnia?

The simple linear model we fitted allows us to propose a relationship between these two variables.

The ANOVA table will help us to analyze the data gathered to determine whether we have enough evidence to conclude that a difference exists somewhere among the treatments, or if the result is due to chance.

The null hypothesis we are testing is that the population mean growth rates for each treatment (infection by various parasite species and no parasite) are all equal. The alternative hypothesis is that there is at least one inequality in the mean growth rates.

10. Calculate the number of observations for each treatment,then print the result.

```
observations <- daphnia %>%
  group_by(parasite.name) %>%
    summarise(observations = length(rep))

observations
```

```
## # A tibble: 4 x 2
##   parasite.name              observations
##   <chr>                             <int>
## 1 Metschnikowia bicuspidata            10
## 2 Pansporella perplexa                 10
## 3 Pasteuria ramosa                     10
## 4 control                              10
```

```
observations = 10
print(observations)
```

```
## [1] 10
```

11. Calculate the number of treatments, then print the result.

```
library(glue)
treatments = length(unique(daphnia$parasite.name))
glue('# of treatments: {treatments}')
```

```
## # of treatments: 4
```

12. Calculate the degrees of freedom from the data.

```
between_group_dof = treatments - 1

within_group_dof = observations - treatments

total_dof = observations - 1

glue('Between group: {between_group_dof}, within group: {within_group_dof}, total: {total_dof}')
```

```
## Between group: 3, within group: 6, total: 9
```

As we will see when we do linear regression, the output from "anova()" will look remarkably similar to that for a regression, and it should. They are both just linear models. In fact, regression analysis and ANOVA are often described as being two sides of the same coin. They are intimately related.

From the results of our analysis (the tabular output of the anova()), We can see that there is indeed evidence that the parasite treatment, comprising four treatments or four levels of manipulation (i. e., a control and three parasites), has produced an effect. The F-value (the value of the test statistic) far exceeds the critical value for our selected alpha value and the and the calculated degrees of freedom.

Notice that the output from "anova()" shows, the between-group variance is large relative to the within-group variance.
This produces a large F-value (test statistic), and thus a small p-value that allows us to
reject the null hypothesis that there are no differences in the population means for the groups.

13. Produce a summary

```
anova(model_grow)
```

```
## Analysis of Variance Table
##
## Response: growth.rate
##           Df Sum Sq Mean Sq F value    Pr(>F)
## parasite   3 3.1379 1.04597  32.325 2.571e-10 ***
## Residuals 36 1.1649 0.03236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's get more information than just the means. This demonstrates that "summarise()" can give us a lot of information if we know how to properly request it using "dplyr".

14. Expand what "summarise()" provides, then print results.

```
sumDat<-daphnia %>%
  group_by(parasite) %>%
    summarise(meanGR = mean(growth.rate))
sumDat
```

```
## # A tibble: 4 x 2
##   parasite meanGR
##   <chr>     <dbl>
## 1 Mb        0.801
## 2 Pp        1.08
## 3 Pr        0.482
## 4 control   1.21
```

Now that we have rejected the null hypothesis that there is no difference in the population means, it would be natural to ask which of the population means are statistically significantly different? One method of doing this was created by the Princeton statistician John W. Tukey. The method is called Tukey's Honestly Significant Difference test (Tukey's HSD). This is one of the things BC&P mean when they write: " . . . if you've ever heard of the 'multiple testing problem', you might be inclined to be careful with those p-values. Ask a [statistician not] a knowledgeable friend for a little guidance if that means nothing to you." We will replace the graph that BC&P prompt you to write a script to produce with a plot from the results of Tukey's HSD. There is nothing wrong with what BC&P want you to do, but we will do something a bit more sophisticated.

As an aside: Tukey developed exploratory data analysis and created many of the methods used in it including the 5-number summary. Please spend a few minutes and Google him to see what other statistical methods he created and, interesting, what commonly used words and phrases he coined.

It should not surprise you that R has Tukey's HSD available (as mentioned above, it is available in both "aov()" and "anova()"). To be able to use Tukey's HSD using 'anova()" for developing the ANOVA table, we need the following libraries

Add these new R libraries:agricolae, and rstatix. Then, we can run Tukey HSD test analysis.

```
# install.packages("agricolae")
# install.packages("rstatix")
library(agricolae)
library(rstatix)
```

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     is_outlier
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

15. Get help

```
?tukey_hsd()
```

Notice what does tukey_hsd() produces and automatically displays. It shows the means that are subtracted (Group 2 - Group 1), the difference in those means (labeled "estimate"), the lower and upper ends of the 95% CI (the default) and the adjusted p-value for Tukey's HSD. The order, top to bottom, in the table is maintained throughout everything that we will see below. That makes interpreting the graphical display that we will ultimately produce easy.

The results of this analysis imply that while the growth rate of infected Daphnia are lower than for noninfected Daphnia, the growth rates are statistically significantly lower than that of the Control group for those Daphnia infected by Metschnikowia bicuspidata

$$p = 4.36 \times 10^{-10}$$

The 95% confidence interval (CI) is $[-0.948, -0.515]$ and

In addition, the decrease in Daphnia growth rates is statistically significantly greater for those Daphnia infected by Pasteuria ramosa relative to
those infected by Pansporella perplexa

$$p = 6.08 \times 10^{-8}$$

The 95% confidence interval (CI) is $[-0.811, -0.377]$

For simplicity in creating a visual display of these results, we will use the core R functions and the installed libraries to redo this analysis. First, we will rerun our ANOVA using "aov()" and display the results.
Then we will redo the post hoc analysis using Tukey's HSD.

In both cases, be sure to compare the results obtained using "anova()" to those obtained using "aov()". You should expect them to be the same.

```
anova(model_grow)
```

```
## Analysis of Variance Table
##
## Response: growth.rate
##           Df Sum Sq Mean Sq F value    Pr(>F)
## parasite   3 3.1379 1.04597  32.325 2.571e-10 ***
## Residuals 36 1.1649 0.03236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov(model_grow)
```

```
## Call:
##    aov(formula = model_grow)
##
## Terms:
##                 parasite Residuals
```

```
## Sum of Squares  3.137907  1.164876
## Deg. of Freedom        3        36
##
## Residual standard error: 0.1798824
## Estimated effects may be unbalanced
```

Now use the function that performs Tukey's HSD, then display the results.

Get help

```
?tukey_hsd(model_grow)
```

Perform the analysis and display the results in tabular form.

```
posthoc_analysis <- tukey_hsd(aov(model_grow))
posthoc_analysis
```

```
## # A tibble: 6 x 9
##   term     group1  group2 null.value estimate conf.low conf.high    p.adj
## * <chr>    <chr>   <chr>       <dbl>    <dbl>    <dbl>     <dbl>    <dbl>
## 1 parasite control Mb              0   -0.413  -0.629    -0.196  5.76e- 5
## 2 parasite control Pp              0   -0.138  -0.354     0.0791 3.34e- 1
## 3 parasite control Pr              0   -0.732  -0.948    -0.515  4.36e-10
## 4 parasite Mb      Pp              0    0.275   0.0585    0.492  8.16e- 3
## 5 parasite Mb      Pr              0   -0.319  -0.536    -0.102  1.82e- 3
## 6 parasite Pp      Pr              0   -0.594  -0.811    -0.377  6.08e- 8
## # i 1 more variable: p.adj.signif <chr>
```

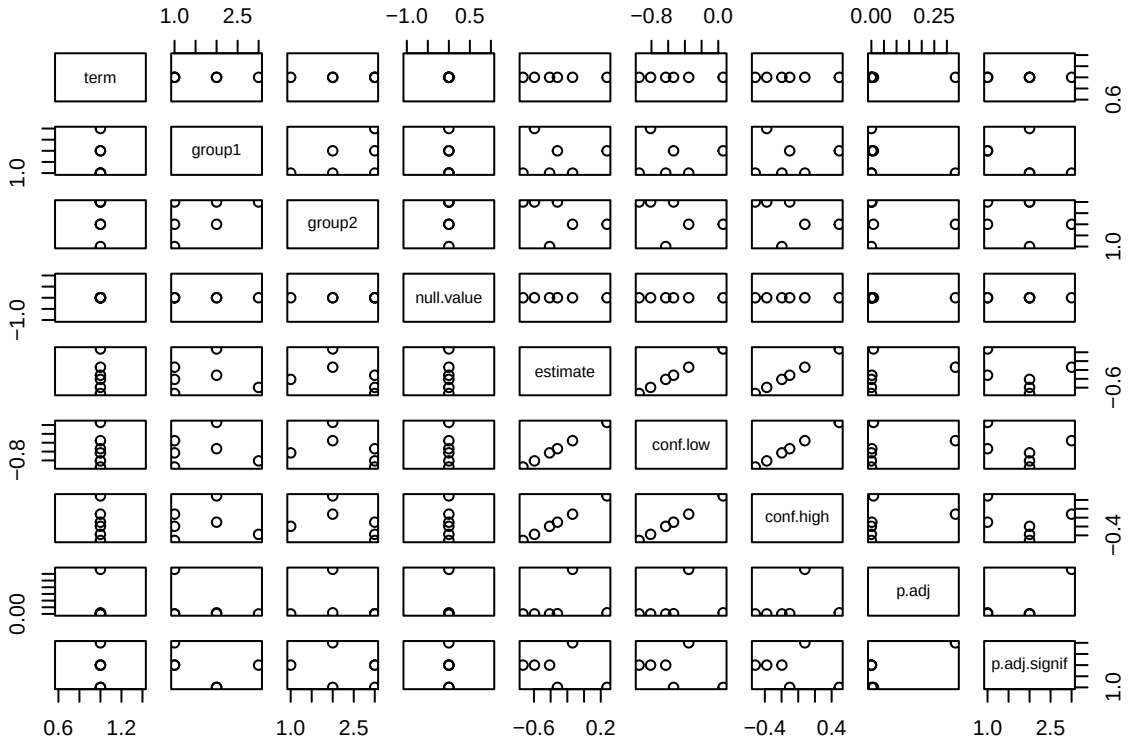
Notice that the results displayed show the pairs of means that are subtracted, the order in which they are subtracted, the resulting difference in the means, the lower and upper limits of the 95% CI (the default), and the adjusted p-value from Tukey's HSD test.

Please check the difference calculation using the table of means that we created above.

Visualization

The following "plot()" function produces a plot of the results from the function TukeyHSD()" using "aov()".

```
plot(posthoc_analysis)
```



The graph displays the 95% CI (the default value) about the mean difference in growth rates for each treatment group pair.

Notice that not every pair of means subtracted is labeled. This is how the person writing the code chose to handle creating a legible graph. But, the subtracted pairs of means in the graph appear from top to bottom in the same order in which they appeared in the output of the script: "posthoc_analysis = TukeyHSD(Test,'parasite')", above. So, we can identify the pairs not shown on the graph.

Those CI's that are different from zero (i.e., those CI's that do not cross the dashed vertical line at 0.0) are the ones of interest. The vertical line at zero indicates no difference in the means.

The graph clearly shows those pairs that are the most different (in this case showing a decrease in growth rate) are those whose mean difference are negative in the post hoc analysis tables. Look at the line for Pr-Pp at the bottom of the graph. You should be able to see the upper (-0.377) and lower (-0.811) limits of the CI that is given in the two versions of the post hoc analysis table; as well as the value for the mean difference (-0.594). Comforting, right?

You can export plots to incorporate into the
assignment you turn in. You may have to experiment with how to do this with the software you are using. We will try to help you with this, but you can always save the plot to your desktop and then print it as a separate page and attach it to your solutions.

If you are using MS WORD, there are at least three ways you can do this:

1. You can export this plot for pasting into a WORD document as a PDF. In the lower right pane of RStudio click "Export" then select "Save Plot as PDF". Select the appropriate for the plot. Select a good size (4.00 x 6.00 may be a good place to start). Check your plot in "Preview". When you think that you have a good plot, select a file name and a directory in which to save it. In WORD, select the Insert tab. In the Text group, click Object. In the Object dialog box, click Create from File. Find the file using Browse, then click Ok.

2. You can also save the plot as a metafile. To do this, select "Export", then "copy to Clipboard", then select "Copy as Metafile". Make the plot have a height of about 600, but you should experiment with this setting to get a good quality graph that clearly displays the relationships in the plot.

3. You can also save your plot as a PDF by writing code. If we want to save the above plot as a PDF we can add the code:

```
pdf("Post-Hoc Analysis Plot")
plot(posthoc_analysis)
dev.off()
```

This will save the PDF to the working directory you set up or are working in. NOTE: You MUST write "dev.off()" at the end of the code or your computer will keeping trying to save as a PDF until R eventually terminates the session; having reached a set time limit.

Now draw the conclusions from the analysis and state them within the context of the problem. Conclusions are written for the indented audience. Write it so that that audience understands what was done and what the conclusions mean in a way meaningful to them.

There are many ways of writing conclusions from statistical analysis of data, but one is as follows.

A one-way analysis of variance was performed to to evaluate if parasitic infections affected growth rates in Daphnia spp. for four different parasite treatment groups, Control (no parasitic infection), Metschnikowia bicuspidata, Pansporella perplexa, and Pasteuria ramosa. Each group consisted of 10 observations.

The mean growth rate +/- SEM for the Control group was

$$1.21 \pm 0.0361$$

This decreased, relative to the Control for Metschnikowia bicuspidata to

$$0.801 \pm 0.0683$$

For Pansporella perplexa:

$$1.08 \pm 0.0568$$

For Pasteuria ramosa:

$$0.482 \pm 0.0613$$

Daphnia growth rates were statistically sign#ificantly different between treatment groups (parasitic infections), assuming an alpha value of 0.05, F = 2.85, versus an F(3,36) = 32.33, p = 2.57e-10.

Post hoc testing analysis using Tukey's Honestly Significant Difference revealed that while the growth rate of infected Daphnia are lower than for noninfected Daphnia.

The growth rates are statistically significantly lower than that of the Control group for those Daphnia infected by Metschnikowia bicuspidata

$$p = 5.76 \times 10^{-5}$$

The 95% confidence interval (CI) is $[-0.629, -0.196]$ and

Pasteuria ramosa:

$$p = 4.36 \times 10^{-10}$$

The 95% confidence interval (CI) is $[-0.948, -0.515]$.

In addition, the decrease in Daphnia growth rates is statistically significantly greater for those Daphnia infected by Pasteuria ramosa relative to
those infected by Pansporella perplexa

$$p = 6.08 \times 10^{-8}$$

The 95% confidence interval (CI) is $[-0.811, -0.377]$

Please make sure that you can calculate the differences shown in the "estimate" and diff" columns of the two post hoc tables.The mean values are given in the display of sumDat1.