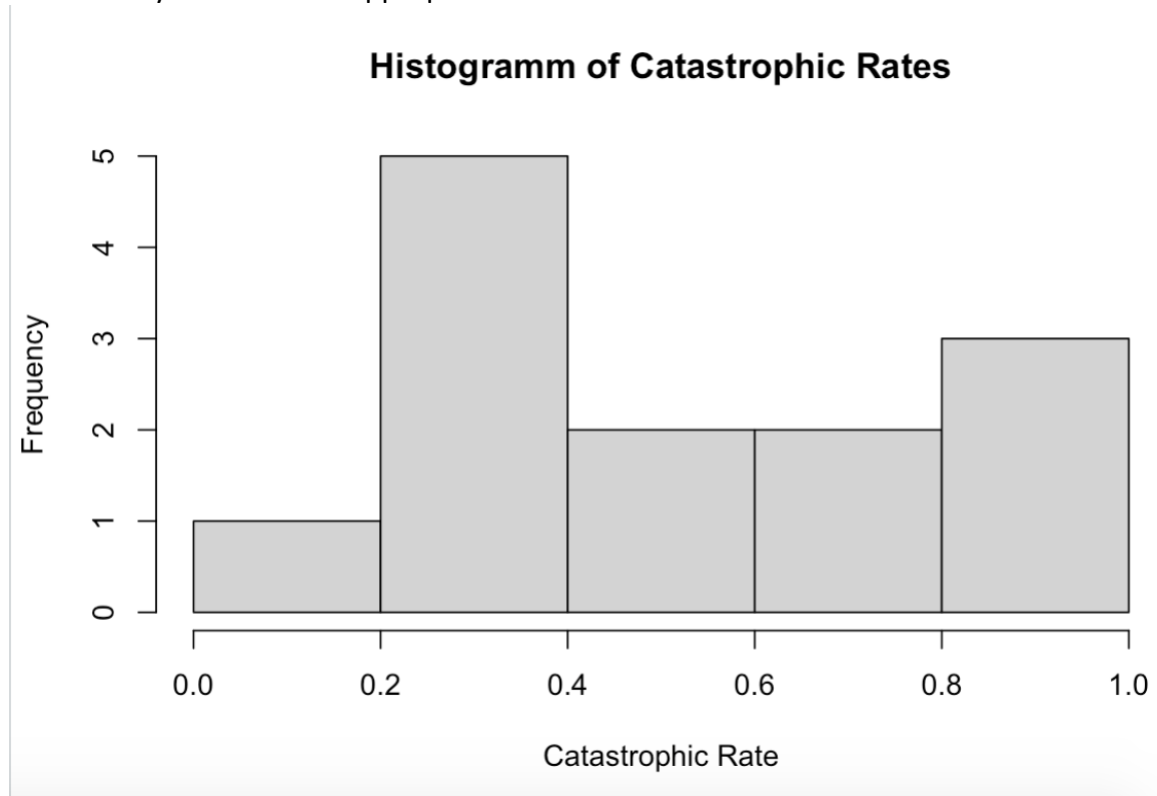


## Analysis of Environmental Data – Using Models 1

Olivia Dinkelacker

**Q1 (1 pt.):** Create a histogram of the salamander reproduction catastrophic rates. Make sure you include an appropriate title and label for the x-axis.



**Q2 (1 pt.):** Conduct a Shapiro-Wilk test of normality of the salamander catastrophic rates. Report the p-value and show the R-code you used to conduct the test.

```
shapiro.test(catrate$cat.rate)  
p-value = 0.04097
```

**Q3 (1 pt.):** What is the null hypothesis for the Shapiro test?  
The salamander catastrophic rate is normally distributed.

**Q4 (1 pt.):** Based on the Shapiro test results, is there strong evidence that the sample came from a non-normally-distributed population?  
Yes, since the p-value is less than 0.05 we assume that the sample is from a non-normally distributed population.

**Q5 (1 pt.):** Show the code you used to conduct the t-test.  
Hint: your answer should only be a single line of code.  
`t.test(catrate$cat.rate, mu = 2/7)`

**Q6 (1 pt.):** State the null hypothesis of the test, in plain nontechnical English.  
The catastrophic rate is not different from the pond late-filling rate.

**Q7 (1 pt.):** Is this a one- or two-tailed test?

It is a two-tailed test, since the catastrophic rate can be different in two ways > greater and less than 2/7.

**Q8 (2 pts.):** What is the p-value from your t-test? Interpret the p-value as a *false-positive rate* using nontechnical English that a non-scientist would understand.

p-value = 0.01193

This value is less than 0.05 which is why we could reject the null hypothesis. But the p-value is a false-positive rate, meaning that we falsely reject the null hypothesis.

**Q9 (1 pt.):** What is the confidence interval for the difference between the null hypothesis and alternative hypothesis means? Did it include zero?

95 percent confidence interval:

0.3526250 0.7261295

The confidence interval does conclude zero, meaning that if we would run the experiment again, we have a good chance of finding no difference between the groups.

**Q10 (2 pts.):** Considering the results from your t-test, did you conclude that there was strong evidence to reject the null hypothesis?

Make sure you justify your answer using the output of the t-test.

On the one hand we have a significant t-test, meaning we could reject the null hypothesis, but it is evened out by the confidence interval, including zero. This means that the output could be different the next time we run the experiment, resulting in **no good evidence to reject the null hypothesis**.

**Q11 (1 pt.):** Show the code you used to conduct the test.

Hint: your answer should only be a single line of code.

```
wilcox.test(catrate$cat.rate, mu = 2/7)
```

**Q12 (1 pt.):** Compare the p-value with the p-value you got from the t-test.

The p-value is much more significant (**p-value = 0.006275**) than the one from the t-test:

p-value = 0.01193.

**Q13 (2 pts.):** Considering the results from your rank sum test, did you conclude that there was strong evidence to reject the null hypothesis?

Make sure you justify your answer using the output of the test.

There is strong evidence to reject the null hypothesis.

**Q14 (1 pt.):** Compare the overall conclusions you could draw from the results of the two tests.

The conclusions vary between the two tests, with the t-test not having a truly significant output. That's probably because the data was not normally distributed and it is not the appropriate test.

**Q15 (1 pt.):** Considering the numerical and graphical data exploration, which test do you think was more appropriate for these data?

The Wilcoxon test is more appropriate since it is used for non-parametric data, and that's what we are dealing with here.

**Q16 (2 pts.):** Show the R-code you used to conduct tests of normality for the flipper lengths of Chinstrap and Adelie penguins.

```
shapiro.test(dat_adelie$flipper_length_mm)
```

```
shapiro.test(dat_chin$flipper_length_mm)
```

**Q17 (2 pts.):** Interpret the test results. Do you conclude that the flipper lengths are normally-distributed for each species? Make sure your answers make reference to the test output.

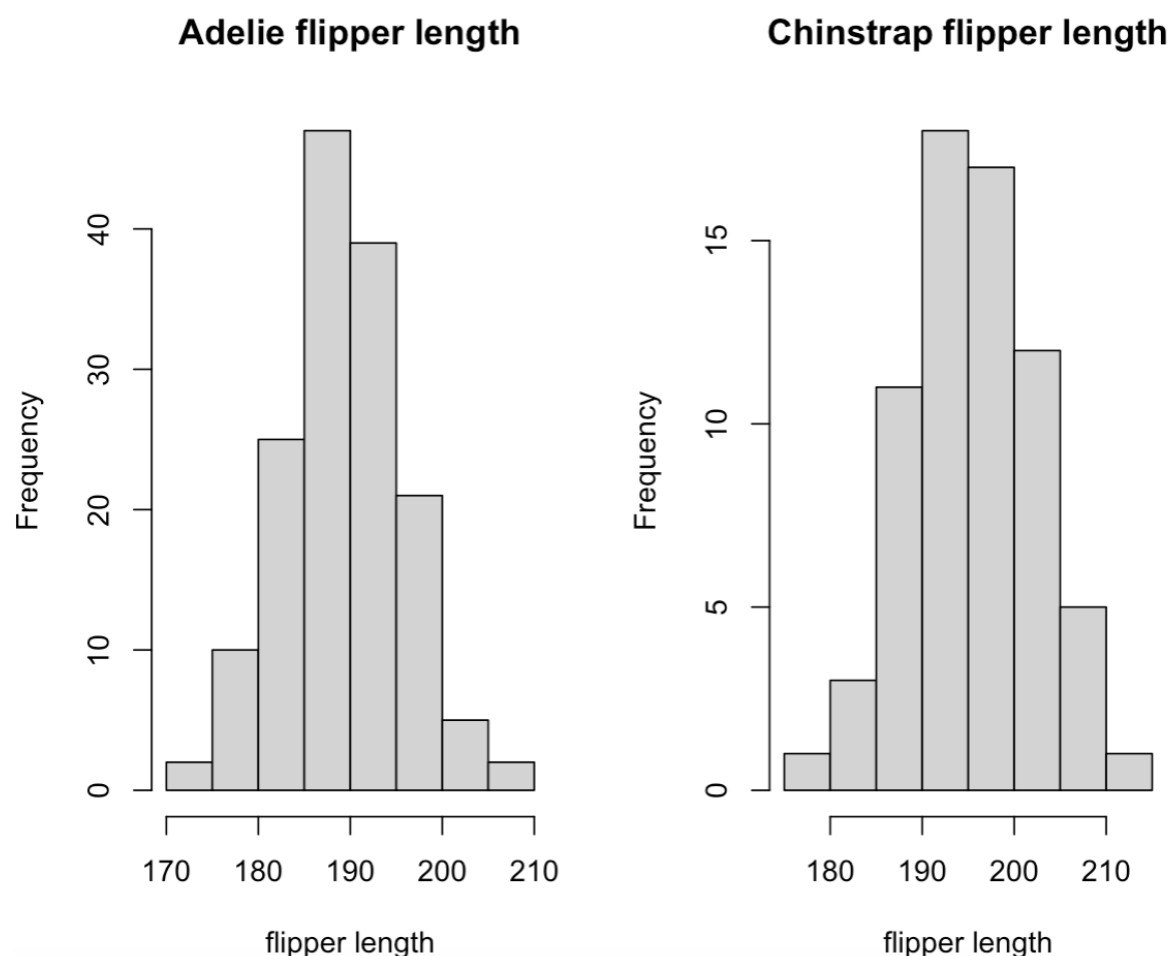
Flipper length is normally distributed in both species, with a p-value higher than 0.05.

Adelie: p-value = 0.72

Chinstrap: p-value = 0.8106

We can accept the null hypothesis that the data is normally distributed.

**Q18 (2 pts.):** Save your figure to a file and include it in your report. Your figure needs to have appropriate dimensions such that the two histograms are not vertically stretched.



**Q19 (2 pts.):** State the alternative hypothesis of the test, in plain nontechnical English.  
Consider whether you used a one- or two- tailed test.  
There is significant difference in the flipper length between Adelie and Chinstrap penguins.

**Q20 (1 pt.):** Include the code you used to conduct the t-test.  
`t.test(dat_adelie$flipper_length_mm, dat_chin$flipper_length_mm)`