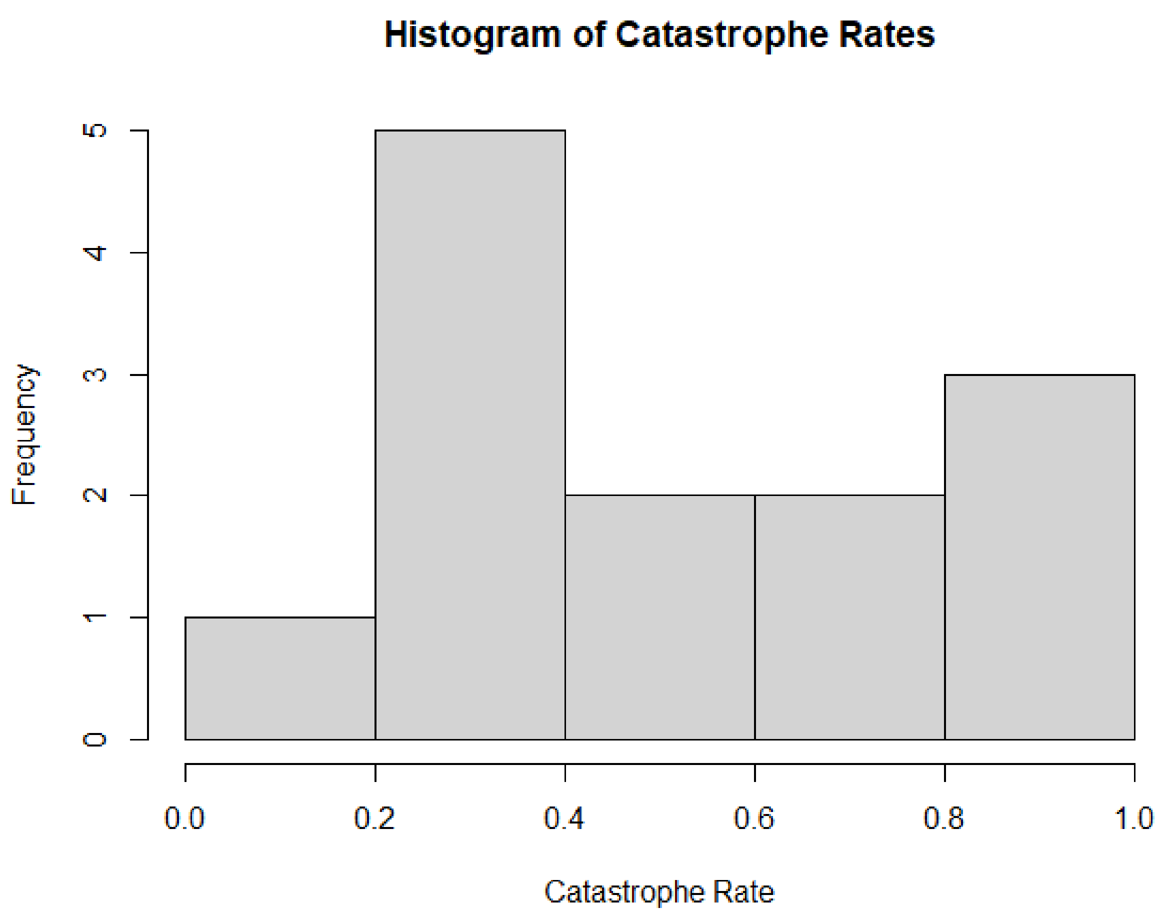
Matthew Jusino

Modeling Assignment 1

- **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)

The p-value is 0.04097

- **Q3 (1 pt.):** What is the null hypothesis for the Shapiro test?

The null hypothesis is that the data for catastrophe rates 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?

Based on the low p-value of 0.04, it is likely that the data came from a non-normally-distributed population and we ought reject the null hypothesis.

- **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 null hypothesis is that the mean of the data, or the average value in the data set, is equal to the late-fill rate.

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

This is a two-sided one-tailed test.

- **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.

The p-value from my t-test is 0.01193. This incredibly small p-value means that it is incredibly unlikely for a false-positive to occur where the sampling would result in a mean of 2/7 for the data set.

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

It did not include zero. The confidence interval means that if the sampling were done repeatedly, 95% of the time the mean value would fall between the lower and upper values of the confidence interval.

- **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.

Yes, there is strong evidence to reject the null hypothesis. Between the incredibly small p-value showing a very low rate of occurrence for mean values of 2/7 in repeat sampling and a 95% confidence interval that does not include the mu value, it is highly unlikely to get a mean of 2/7, and as such, it is appropriate to reject the null hypothesis that the mean of the cat rate is not different than the late-fill rate.

- **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 for the Wilcoxon test is 0.006275, significantly smaller than the p-value of the t-test.

- **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

Still yes. In fact, with the even lower p-value for the Wilcoxon test, I would say that there is even stronger evidence to reject the null hypothesis based off this p-value of 0.006

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

I would say that while I felt both tests gave adequate information to reject the null hypothesis, the addition of the confidence interval in the student's t-test gave even more evidence towards a conclusion.

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

I think while the student's t-test gave more information to draw conclusions from, the Wilcoxon test gave a higher accuracy in the results, as the p-value was significantly lower, leading to better confidence in the decision to reject the nullhypothesiss.

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

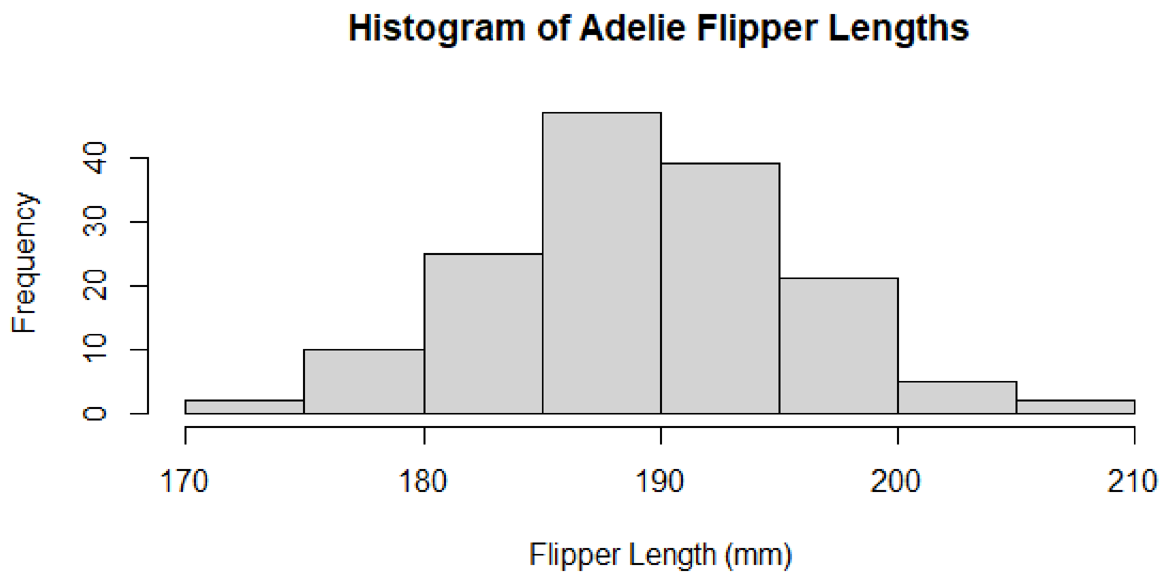
shapiro.test(adelie$flipper_length_mm)

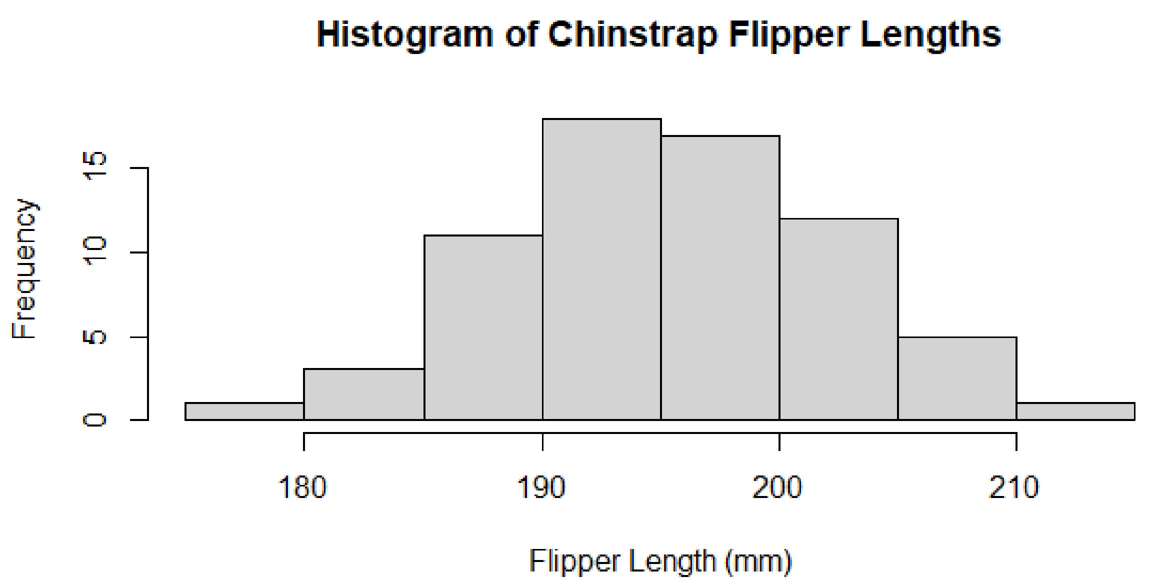
shapiro.test(chinstrap$flipper_length_mm)

- **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.

The Adelie flipper lengths had a p-value of 0.72 and the Chinstrap flipper lengths had a p-value of 0.81. Both of these p-values are quite large, leading to the conclusion that the flipper lengths of the two species are essentially 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.
    - Hint: Check out the `width` and `height` arguments.
    - Hint: Remember the `par()` function? Which argument did we use to include multiple plots in the same figure?





- **Q19 (2 pts.):** State the alternative hypothesis of the test, in plain nontechnical English.
  - Consider whether you used a one- or two- tailed test.

The alternative hypothesis is that Adelie and Chinstrap penguins have different average flipper lengths. It is a two-tailed test, as it is testing for difference, so it is testing both greater and less than between the two sets.

- **Q20 (1 pt.):** Include the code you used to conduct the t-test.
  - Hint: your answer should only be a single line of code.

```
t.test(adelie$flipper_length_mm, chinstrap$flipper_length_mm)
```