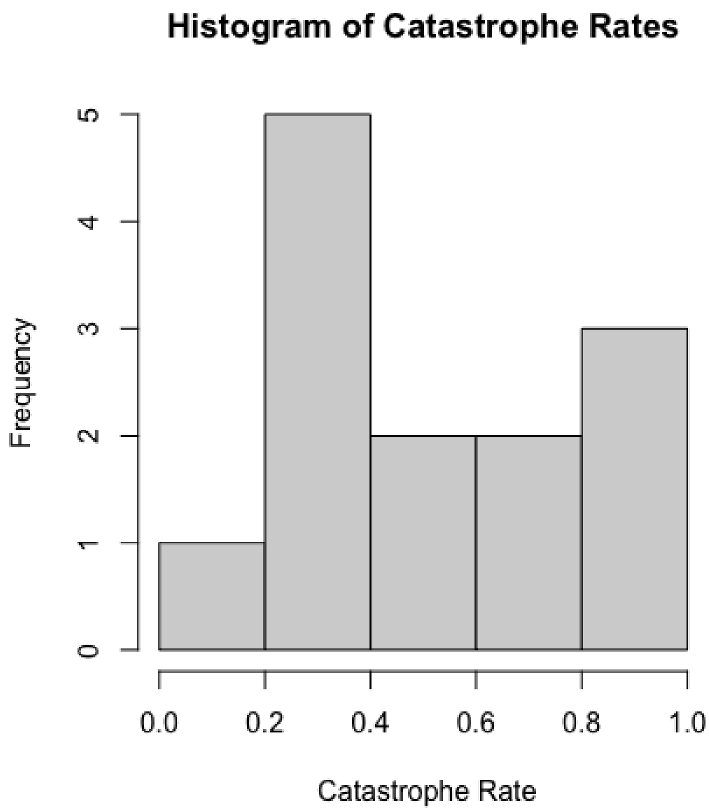
Kaitlyn Valentinetti

November 13, 2022

ECo 602: Michael France Nelson

Lecture Assignment: Using Models I

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 null hypothesis for the Shapiro test is the catastrophe rate data are 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?
There is strong evidence that the sample came from a non-normally distributed population because the p-value is less than 0.05 (it's actually 0.04097).

Q5 (1 pt.): Show the code you used to conduct the t-test.
t.test(x = catrate$cat.rate, mu = 0.2857143)

Q6 (1 pt.): State the null hypothesis of the test, in plain nontechnical English.
There is no difference between the observed and expected catastrophe rate.

Q7 (1 pt.): Is this a one- or two-tailed test?
This is a two-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 is 0.01193. This means that the probability of rejecting a true null hypothesis, also known as the false positive rate, is 1.2%. We consider that you can reject the null hypothesis if the p-value is less than 0.05.

Q9 (1 pt.): What is the confidence interval for the difference between the null hypothesis and alternative hypothesis means? Did it include zero?
The confidence interval is: 0.3526250 to 0.7261295. The interval therefore does include zero.

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.
Considering the results from the t-test, we might be tempted to conclude there is strong evidence to reject the null hypothesis since the p-value is less than 0.05; however, the fact that the

confidence interval is around zero suggests that there is not a significant difference between the observed and expected catastrophe 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 of the Wilcoxon test is 0.006275, whereas the p-value from the t-test is 0.01193. The Wilcoxon test p-value is much more statistically significant in terms of rejecting the null hypothesis.

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 because the p-value is so low.

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

You can draw the conclusion from Wilcoxon test that the observed values will be different than the expected catastrophe rate (rejecting the null hypothesis), but you cannot conclude this from the t-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 for this data because it is non-normally distributed and coming from a relatively small sample.

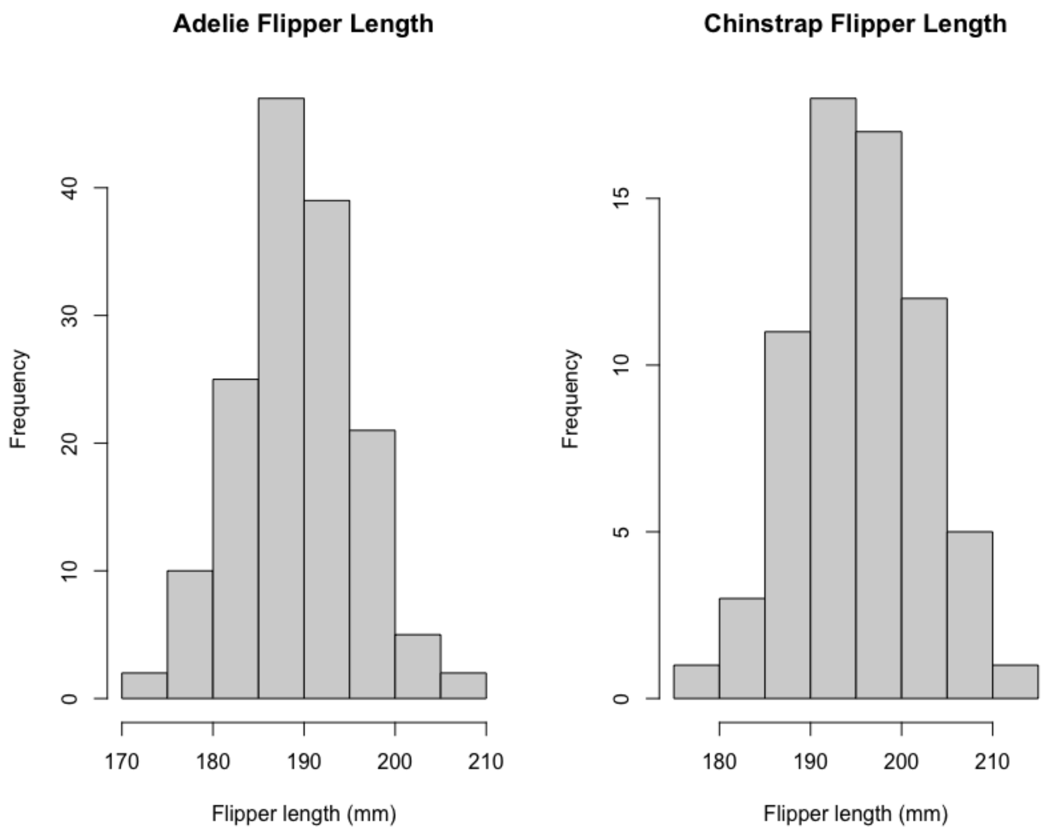
Q16 (2 pts.): Show the R-code you used to conduct tests of normality for the flipper lengths of Chinstrap and Adelie penguins.

```
shapiro.test(x = dat_adelie$flipper_length_mm)
shapiro.test(x = dat_chinstrap$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 refer to the test output.

Since Adelie penguin's p-value is 0.72, and Chinstrap penguin's p-value is 0.8106, the flipper lengths are normally distributed in each population. A p-value greater than or equal to 0.05 indicates a normally distributed population.

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.

The alternative hypothesis of the test is there is a significant difference in the flipper length between Adelie and Chinstrap penguin species.

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(flipper_length_mm ~ species, data = penguin_dat)
```