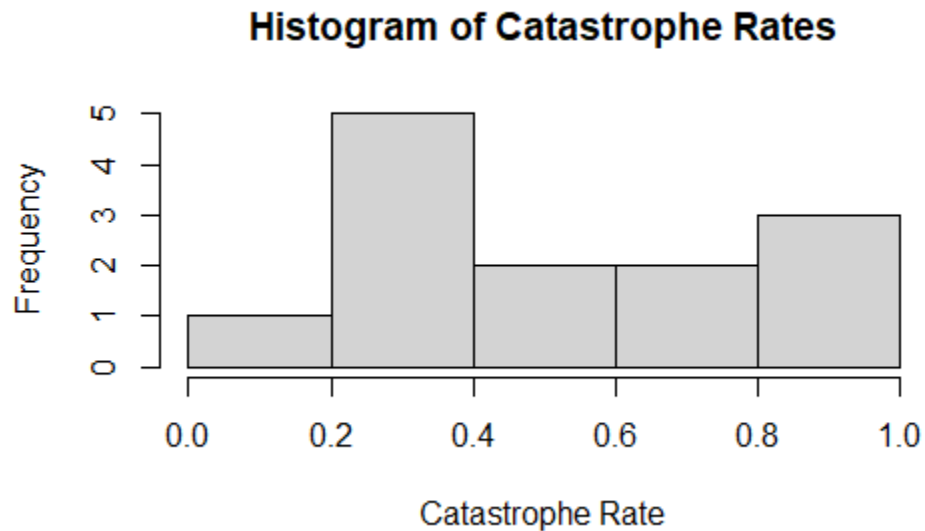


Lina Clifford  
ECO 602 – Analysis of Environmental Data  
Lecture Assignment: Using Models 1  
Due 11/13/2022

**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)
```

Shapiro-Wilk normality test

data: catrate\$cat.rate

W = 0.86202, p-value = 0.04097

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

The null hypothesis for the Shapiro test is that the 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?

Based on the 0.04097 p-value output from the Shapiro test, there is strong evidence that the sample came 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, alternative = "two.sided", mu = 2/7)
```

**Q6 (1 pt.): State the null hypothesis of the test, in plain nontechnical English.**

The null hypothesis of the test is that the late pond filling rate is not significantly different from the 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.**

p-value = 0.005966

This p-value is less than 0.05, which means the means there was a significant difference in catastrophe rate and the late-pond filling rate. There is a 0.5% likelihood that we have a false positive in detecting a significant difference.

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

95% confidence interval for this t-test is 0.3866123 and does NOT include 0.

**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.**

From this t-test's p-value, 0.005966, there is strong evidence to reject the null hypothesis. This is because the p-value is less than 0.05.

**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 from the second t-test was 0.005966 and the p-value from the Wilcox test was 0.006275. The Wilcox test p-value is slightly larger.

**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 since the Wilcox test p-value was less than 0.05.

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

From the Wilcox test, we can conclude that there is a significant difference between the catastrophe rate and the late-pond filling rate. Since our sample is not from a normally distributed population, we cannot draw much concrete conclusions from the t-test as the t-test assumes normality.

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

The Wilcox test is likely more appropriate for these data because they failed the Shapiro normality test. Since the Wilcox test is non-parametric and does not assume the data to be normally distributed, it is the better test for these data.

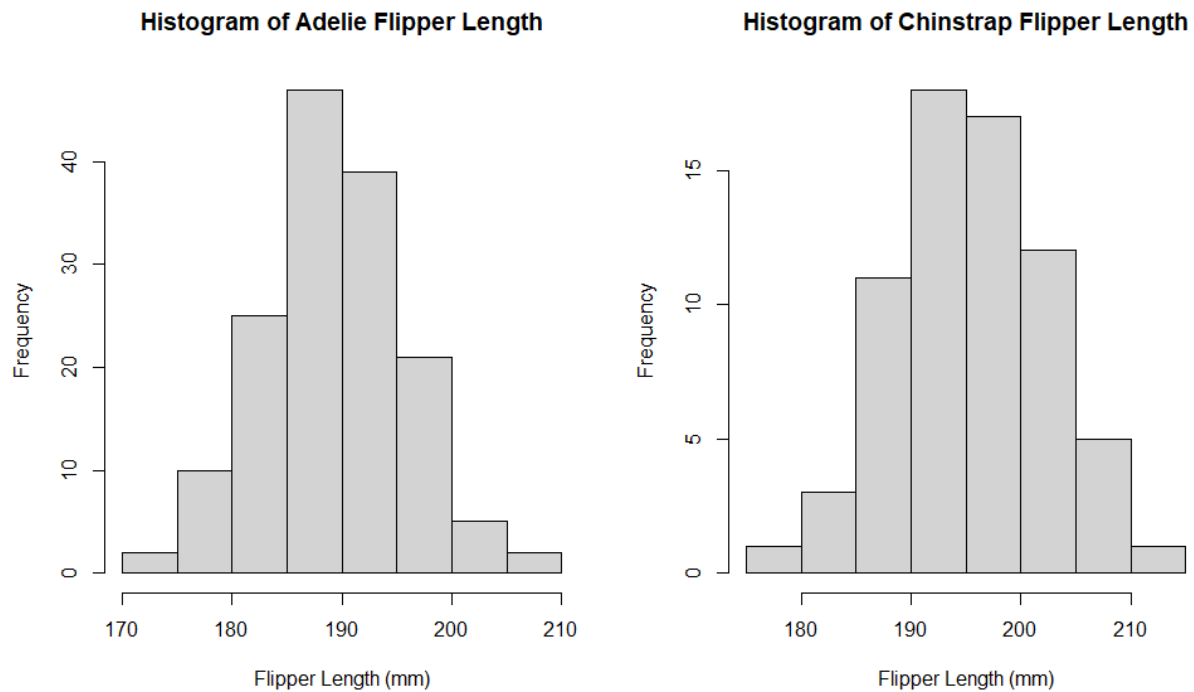
**Q16 (2 pts.): Show the R-code you used to conduct tests of normality for the flipper lengths of Chinstrap and Adelie penguins.**

```
shapiro.test(dat_adelie$flipper_length_mm)
shapiro.test(dat_chinstrap$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 resulting p-values from the Shapiro tests for both species is much, much greater than 0.05. This tells me that the flipper lengths are normally distributed for each species. The p-value for Adelie was 0.72 and the p-value for Chinstrap was 0.8106.

**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 there is a significant difference between the mean flipper length of Adelie and Chinstrap penguin species. This would be a two tailed test.

**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(dat_adelie$flipper_length_mm, y =  
dat_chinstrap$flipper_length_mm, alternative = "two.sided", mu = 0)
```