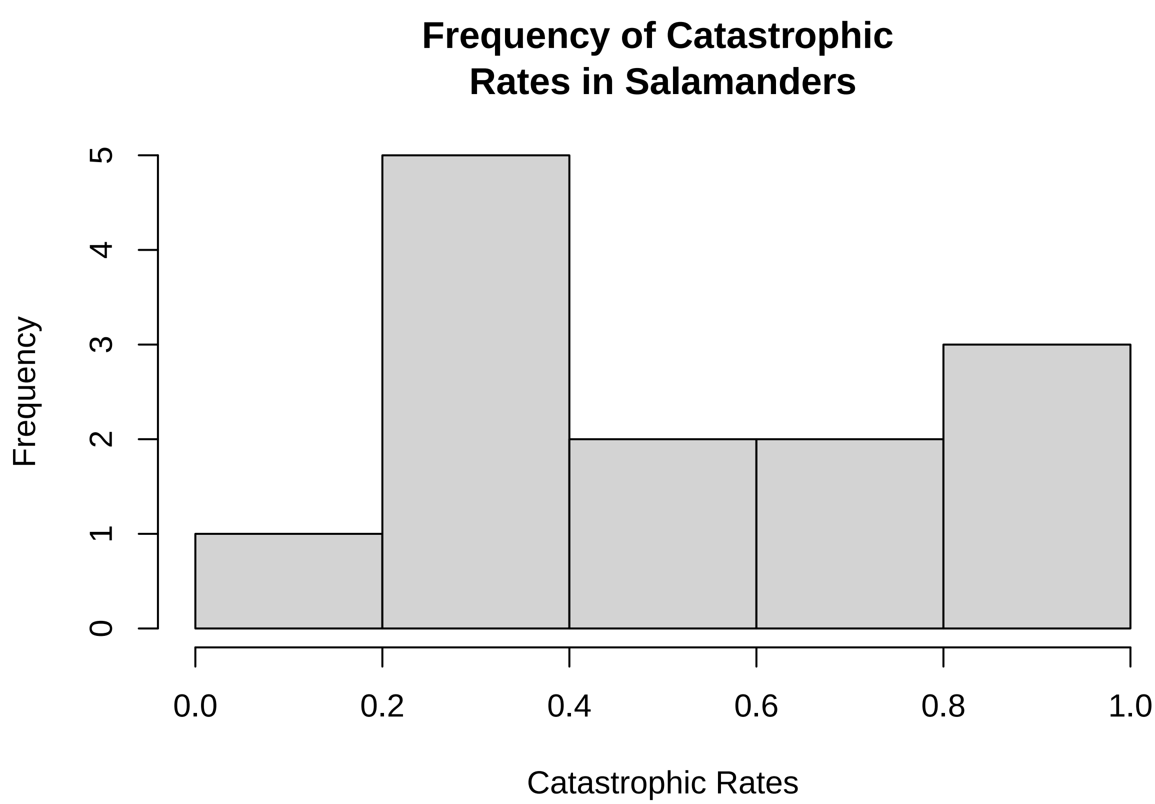
Using Models 1

Ethan Rutledge

11/2/2021

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.

```
hist(catrate$cat.rate, ylab = "Frequency",
     xlab = "Catastrophic Rates",
     main = "Frequency of Catastrophic \nRates in Salamanders")
```



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. Pvalue = 0.04097

```
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-Wilk test is: "The data were sampled from a normally-distributed population".

Q4 (1 pt.): Based on the Shapiro test results, is there strong evidence that the sample came from a non-normally-distributed population? Yes there is evidence that the sample came from a non-normally distributed population.

Q5 (1 pt.): Conduct a one-sample t-test of the alternative hypothesis that the catastrophic rate is different from the pond late-filling rate. Show the code you used to conduct the t-test

```r
t.test(catrate$cat.rate,
       alternative = c("two.sided"),
       mu =  2/7)
```

```
##
##  One Sample t-test
##
## data:  catrate$cat.rate
## t = 2.9595, df = 12, p-value = 0.01193
## alternative hypothesis: true mean is not equal to 0.2857143
## 95 percent confidence interval:
##  0.3526250 0.7261295
## sample estimates:
## mean of x
## 0.5393773
```

Q6 (1 pt.): State the null hypothesis of the test, in plain nontechnical English. The null hypothesis is that there is no difference between the pond late-filling rate and the catastrophic rate.

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

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, we could sample the population many times and we would expect to find that the pond-late filling rate is equal to the catastrophic rate only 1 in 100 times.

Q9 (1 pt.): What is the confidence interval for the difference between the null hypothesis and alternative hypothesis means? Did it include zero? 0.3526250 0.7261295, it does not 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? There is strong evidence to reject the null hypothesis, there is a difference between the pond late-filling rate and the catastrophic rate. The mean of the null hypothesis falls outside of the confidence interval of the alternative hypothesis. The p-value is $<.05$.

Q11 (1 pt.): Conduct a one-sample Wilcoxon rank sum test of the alternative hypothesis that the catastrophic rate is different from the pond late-filling rate. Show the code you used to conduct the test.

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

```
## Warning in wilcox.test.default(catrate$cat.rate, mu = 2/7): cannot compute exact
## p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  catrate$cat.rate
## V = 85, p-value = 0.006275
## alternative hypothesis: true location is not equal to 0.2857143
```

Q12 (1 pt.): Compare the p-value with the p-value you got from the t-test. The p-value for the wilcox test is smaller than the p-value for the Ttest. Probably because its a more suitable test for non normal distributions.

Q13 (2 pts.): Considering the results from your rank sum test, did you conclude that there was strong evidence to reject the null hypothesis? There is strong evidence to reject the null hypothesis, the p-value is $<.05$

Q14 (1 pt.): Compare the overall conclusions you could draw from the results of the two tests. Both tests reached rejected the null hypothesis, though the wilcox test was stronger.

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 more suitable for these data because both the histogram and the shapiro test show that the salamander catastophe data is not normally distributed.

Q16 (1 pt.):Show the R-code you used to conduct tests of normality for the flipper lengths of Chinstrap and Adelie penguins.

```r
require(palmerpenguins)
```

```
## Loading required package: palmerpenguins
```

```r
penguin_dat = droplevels(subset(penguins, species != "Gentoo"))
```

```r
dat_adelie = subset(penguin_dat, species == "Adelie")
dat_chinstrap = subset(penguin_dat, species == "Chinstrap")

shapiro.test(dat_adelie$flipper_length_mm)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dat_adelie$flipper_length_mm
## W = 0.99339, p-value = 0.72
```

```r
shapiro.test(dat_chinstrap$flipper_length_mm)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dat_chinstrap$flipper_length_mm
## W = 0.98891, p-value = 0.8106
```

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. For both the Adelie and the Chinstrap penguins the data were sampled from a normally-distributed population. Both p-values were $>.05$

Q18 (2 pts.): Create a single figure consisting of histograms of flipper lengths of Adelie and Chinstrap penguins. Your figure should have two histograms arranged in one row 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.

```
png(filename = here("images", "Flipper_Length_Hist.png"),
    width = 1800, height = 700, res = 180)

par(mfrow = c(1, 2))
hist(dat_adelie$flipper_length_mm,
     main = " Flipper Lengths of \nAdelie Penguins",
     xlab = "Flipper Length")
hist(dat_chinstrap$flipper_length_mm,
     main = " Flipper Lengths of \nChinstrap Penguins",
     xlab = "Flipper Length")

dev.off()
```

```
## pdf
##   2
```

Q19 (2 pts.): State the alternative hypothesis of the test, in plain nontechnical English. Consider whether you used a one- or two- tailed test.Conduct a two-sample t-test of the alternative hypothesis that the Adelie penguins have different flipper lengths than Chinstrap penguins. I conducted a two-tailed test.The alternative hypothesis is that there is a difference in mean flipper length between Adelie and chinstrap penguins.

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)
```

```
##
##  Welch Two Sample t-test
##
## data:  flipper_length_mm by species
## t = -5.7804, df = 119.68, p-value = 6.049e-08
## alternative hypothesis: true difference in means between group Adelie and group Chinstrap is not equa
## 95 percent confidence interval:
##  -7.880530 -3.859244
## sample estimates:
##    mean in group Adelie mean in group Chinstrap
##                189.9536                195.8235
```