

UsingModels1.R

Feipeng Huang

2022-11-10

```
catrate =  
read.csv("/Users/stonehuang/Documents/environmental_data/data/catrate.csv")  
head(catrate)
```

```
##   pond success years  cat.rate  
## 1    2      5     7 0.2857143  
## 2    3      5     7 0.2857143  
## 3    4      6     7 0.1428571  
## 4    5      4     7 0.4285714  
## 5    6      0     7 1.0000000  
## 6    7      1     4 0.7500000
```

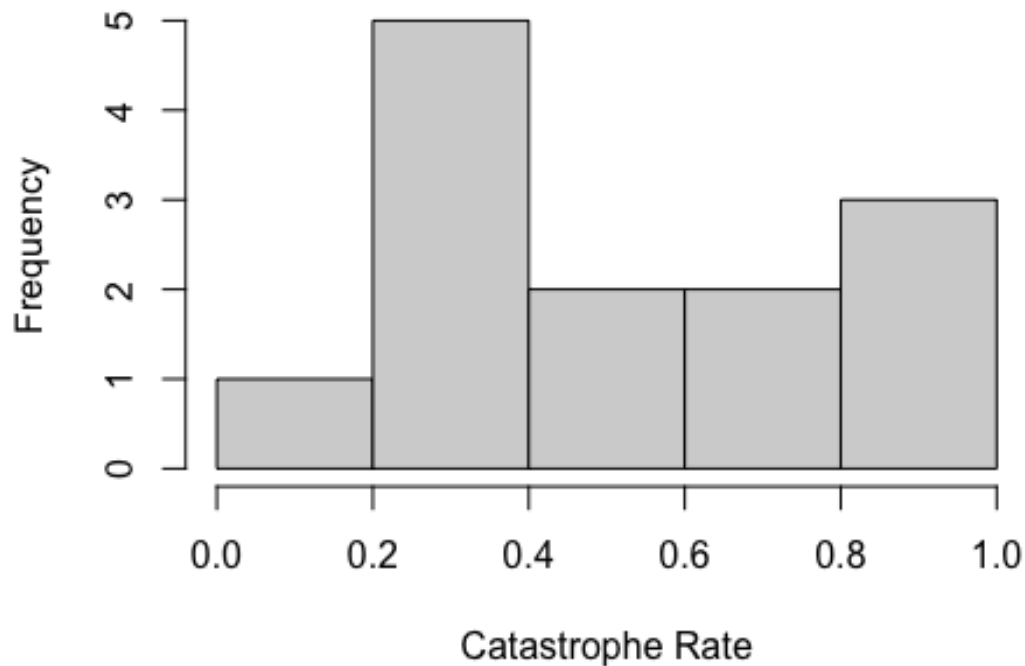
```
summary(catrate)
```

```
##      pond      success      years      cat.rate  
## Min.   : 2   Min.   :0.000   Min.   :1.000   Min.   :0.1429  
## 1st Qu.: 5   1st Qu.:1.000   1st Qu.:3.000   1st Qu.:0.2857  
## Median : 8   Median :2.000   Median :4.000   Median :0.4286  
## Mean   : 8   Mean   :2.538   Mean   :4.692   Mean   :0.5394  
## 3rd Qu.:11   3rd Qu.:5.000   3rd Qu.:7.000   3rd Qu.:0.7500  
## Max.   :14   Max.   :6.000   Max.   :7.000   Max.   :1.0000
```

#Q1.

```
hist(catrate$cat.rate, xlab = "Catastrophe Rate", main = "Histogram of  
Catastrophe Rates")
```

Histogram of Catastrophe Rates



#Q2. p-value = 0.04097

```
shapiro.test(catrate$cat.rate)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  catrate$cat.rate
```

```
## W = 0.86202, p-value = 0.04097
```

#Q3. The null hypothesis for the Shapiro-Wilk test is that the data were sampled from a normally-distributed population.

#Q4. There is strong evidence that the sample came from a non-normally-distributed population.

#Q5.

```
t.test(catrate$cat.rate, 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. The catastrophic rate is equal to the pond late-filling rate.
#Q7. This is a two-tailed test.
#Q8. p-value = 0.01193. Even if the p-value is very small and we think there
is no pattern, there is still a small chance that a pattern exists.
#Q9. 0.3526250 0.7261295, it does not include 0
#Q10. I conclude that there is strong evidence to reject the null hypothesis
because the p-value is lower than 0.05.
#Q11.
wilcox.test(catrate$cat.rate, mu = 2 / 7, exact=FALSE)

##
## 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. The p-value is 0.006275, lower than the p-value from the t-test.
#Q13. I conclude that there is strong evidence to reject the null hypothesis
because the p-value is lower than 0.05.
#Q14. The overall conclusions I could draw from the results of the two tests
are the same.
#Q15. The Wilcoxon Rank Sum Test is more appropriate because the data are not
normally-distributed.
#Q16.
require(palmerpenguins)

## Loading required package: palmerpenguins

penguin_dat = droplevels(subset(penguins, species != "Gentoo"))
dat_adelie = subset(penguin_dat, species == "Adelie")
shapiro.test(dat_adelie$flipper_length_mm)

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

dat_Chinstrap = subset(penguin_dat, species == "Chinstrap")
shapiro.test(dat_Chinstrap$flipper_length_mm)

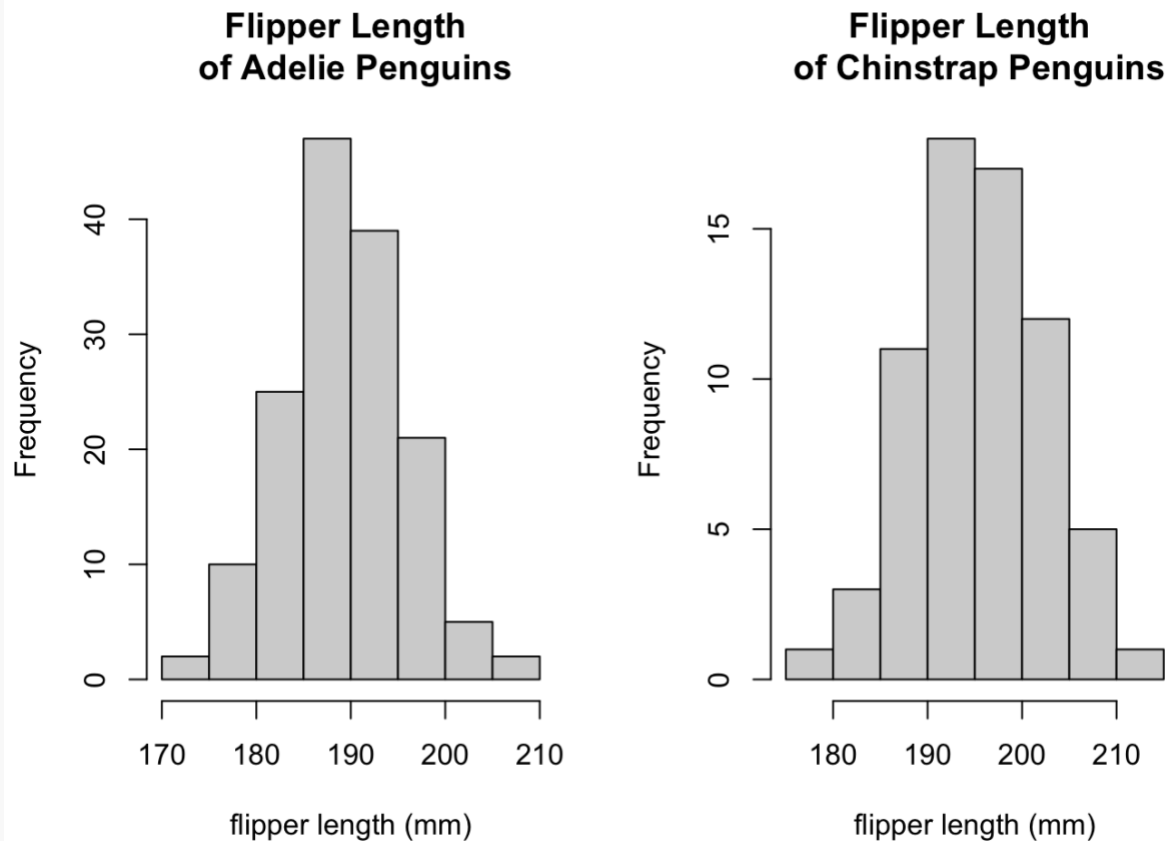
##
## Shapiro-Wilk normality test
##
## data:  dat_Chinstrap$flipper_length_mm
## W = 0.98891, p-value = 0.8106

```

#Q17. I conclude that the flipper lengths are normally-distributed for both species because the p-values (0.72 for Adelie, 0.8106 for Chinstrap) are higher than 0.05. We fail to reject the null hypothesis.

#Q18.

```
par(mfrow = c(1, 2))
hist(dat_adelie$flipper_length_mm, xlab = "flipper length (mm)", main =
"Flipper Length \n of Adelie Penguins")
hist(dat_chinstrap$flipper_length_mm, xlab = "flipper length (mm)", main =
"Flipper Length \n of Chinstrap Penguins")
```



#Q19. Adelie and Chinstrap penguins have different flipper lengths. (two-tailed test)

#Q20.

```
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 is not equal to 0
## 95 percent confidence interval:
## -7.880530 -3.859244
```

