Migration Case Study: Patterns in sicklefin redhorse migration

**Introduction**

Sicklefin redhorse are a species of sucker in the genus *Moxostoma* that are a candidate for the endangered species list. Various agencies and entities partnered together to conduct conservation work to protect, restore, and monitor sicklefin redhorse populations throughout their range. As part of this work, biologists from the Georgia Department of Natural Resources, Tennessee Valley Authority, the U.S. Department of Natural Resources, and other groups collected sicklefin redhorse from Brasstown Creek in north Georgia. Captured individuals were tagged with passive integrated transponder (PIT) tags during their spawning migration and released back into the stream. A PIT tag detection antenna was deployed at a fixed site along the stream bottom to record any tagged sicklefin redhorse that passed over the antenna. This data allowed biologists to determine the timing of migration and the duration of migration in the spawning site. Biologists also recorded the sex of individuals and their total length. The data set below provides detection data from 2017 and 2018.

The data set can be loaded into R with the code below. Users will need to set the working directory to the location of the data set on their computer.

#Set the working directory appropriately.
setwd("C:/Users/davisjg/Desktop/FMN Work")
Migration <- read.csv("SFR Movement Data.csv")

The data set contains the following columns of data:

**Individual** – A unique number assigned to each individual fish

**First Date** – The first date that an individual was detected by the antenna, representing when the individual migrated into the spawning grounds.

**Last Date** – The last date that an individual was detected by the antenna, representing when the individual migrated back downstream.

**Year** – The year of the detection, being either 2017 or 2018.

**Residence Time** – The number of days between the first date of detection and the last date of detection, representing the time of residency in the spawning grounds.

**Total Length** – The size (mm) of the sicklefin redhorse as measured from the tip of the snout to the tip of its caudal fin (or tail).

**Sex** – The biological sex of the individual (male or female) based upon the expression of spawning tubercles on the fins.

### Assignment: You have been provided with the results of an analysis of the data set. Several hypotheses are presented, and you should interpret the results of the analysis relative to the presented hypotheses. You will evaluate the results in the context of what has been learned in the case study thus far.

### Hypothesis: There is no difference in residence time between years.

To compare the mean residence time of sicklefin redhorse in the spawning grounds in 2017 and 2018, an appropriate test to compare means between two groups in needed.

First, the assumption that the residence time follows a normal distribution should be tested. This can be done by constructing a histogram and Q-Q plot of residence time and using a Shapiro-Wilk test to determine if the null hypothesis of the data having a normal distribution is supported.

hist(Migration$Residence.Time)



*Question*: Does the histogram show a normal distribution of the data? Describe the shape of the distribution?

*Answer:*

library(car)
qqPlot(Migration$Residence.Time)



*Question:* If a Q-Q plot has values that are near the line of best fit and within the shaded region, then there is evidence that the data fits a normal distribution? Does the Q-Q plot indicate a normal distribution for residence time?

*Answer:*

shapiro.test(Migration$Residence.Time)

##
## Shapiro-Wilk normality test
##
## data: Migration$Residence.Time
## W = 0.90586, p-value = 3.121e-08

*Question:* The null hypothesis of the Shapiro-Wilk test is as follows:

H0: The data fits a normal distribution.

Based upon the results of the test, is the data normally distributed?

*Answer:*

As a result of the data not fitting a normal distribution, a non-parametric test is appropriate. The Mann-Whitney U test is a non-parametric alternative to the more typical t-test. The Mann-Whitney U test determines whether the two groups have a similar distribution or compares the medians between two groups. The following statistical hypotheses are tested:

H0: There is no significant difference between the median residence time between years.

HA: There is a significant difference between the median residence time between years.

Migration$Year <- factor(Migration$Year)
wilcox.test(Migration$Residence.Time~Migration$Year)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Migration$Residence.Time by Migration$Year
## W = 1684, p-value = 4.989e-05
## alternative hypothesis: true location shift is not equal to 0

*Question:* Based upon the results of the Mann-Whitney U test, does residence time of sicklefin redhorse differ between 2017 and 2018?

*Answer:*

The outcome of this analysis can be illustrated by a variety of plots, including a box plot, violin plot, strip chart, and histograms. Use the code below to generate each of these plots.

boxplot(Migration$Residence.Time~Migration$Year,
 main = "Comparison of Residence Time Between Years",
 xlab = "Year",
 ylab = "Residence Time (Days)")

library(ggplot2)
ggplot(Migration, aes(x=Year, y=Residence.Time, fill = Year)) +
 geom\_violin() +
 xlab("Year") + ylab("Residence Time (Days)") +
 theme\_classic()+scale\_fill\_manual(values=c("#FFB531","#BC211A"))+
 stat\_summary(fun.y=mean, geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)



The box plots shows the median as a horizontal line inside the box. The edges of the boxes represent the 25th and 75th percentiles.



A violin plot depicts distributions of data for groups using density curves. The width of each curve corresponds with the approximate frequency of data points in each region.

ggplot(Migration, aes(x = Year, y = Residence.Time)) +
 geom\_point(color = "firebrick", size = 3, shape = 1) +
 stat\_summary(fun.data = mean\_cl\_normal, geom = "errorbar",
 width = 0.1, position=position\_nudge(x = 0.15))+
 stat\_summary(fun.y = mean, geom = "point", color = "firebrick", size = 3, position=position\_nudge(x = 0.15)) +
 labs(x = "Year", y = "Residence Time (Days)") +
 theme\_classic()



A strip chart is valuable because it also shows the location of all of the data points. This strip chart displays the mean values and standard error of the mean.

library(dplyr)
library(hrbrthemes)
p <- Migration %>%
 ggplot( aes(x=Residence.Time, fill=Year)) +
 geom\_histogram( color="#e9ecef", alpha=0.5, position = 'identity') +
 scale\_fill\_manual(values=c("#69b3a2", "#404080")) +
 theme\_ipsum() +
 labs(x = "Year", y = "Residence Time (Days)") +
 theme\_classic()
p



This histogram shows the distributions of each group overlapping each other. Thus, the histogram allows for a comparison of similarity of distributions.

*Question:* Based upon the results of the analysis and the visual representation of the data from the graphs, summarize the results of the analysis.

*Answer:*

### Hypothesis: There is no difference in residence time between sexes.

Similar to the previous analysis, the hypothesis that residence time may differ between males and females can be tested. Because residence time is not normally distributed, the Mann-Whitney U test is used again.

Migration$Sex <- factor(Migration$Sex)
wilcox.test(Migration$Residence.Time~Migration$Sex)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Migration$Residence.Time by Migration$Sex
## W = 308.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0

*Question:* What is the null and alternative hypothesis?

*Answer:*

*Question:* Based upon the results of the analysis, what conclusion can be drawn?

*Answer:*

Again, the outcome of this analysis can be illustrated by a box plot, violin plot, strip chart, and histogram. Construct the graphs using the following R code:

boxplot(Migration$Residence.Time~Migration$Sex,
 main = "Comparison of Residence Time Between Sex",
 xlab = "Sex",
 ylab = "Residence Time (Days)")

library(ggplot2)
ggplot(Migration, aes(x=Sex, y=Residence.Time, fill = Sex)) +
 geom\_violin() +
 xlab("Sex") + ylab("Residence Time (Days)") +
 theme\_classic()+scale\_fill\_manual(values=c("pink","blue"))+
 stat\_summary(fun.y=mean, geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)



ggplot(Migration, aes(x = Sex, y = Residence.Time)) +
 geom\_point(color = "firebrick", size = 3, shape = 1) +
 stat\_summary(fun.data = mean\_cl\_normal, geom = "errorbar",
 width = 0.1, position=position\_nudge(x = 0.10))+
 stat\_summary(fun.y = mean, geom = "point", color = "firebrick", size = 3, position=position\_nudge(x = 0.10)) +
 labs(x = "Sex", y = "Residence Time (Days)") +
 theme\_classic()

library(dplyr)
library(hrbrthemes)
p <- Migration %>%
 ggplot( aes(x=Residence.Time, fill=Sex)) +
 geom\_histogram( color="#e9ecef", alpha=0.5, position = 'identity') +
 scale\_fill\_manual(values=c("pink", "#404080")) +
 theme\_ipsum() +
 labs(x = "Sex", y = "Residence Time (Days)") +
 theme\_classic()
p





*Question:* Based upon the results of the analysis and the visual representation of the data from the graphs, summarize the results of the analysis.

*Answer:*

### Hypothesis: There is no relationship between residence time and fish size.

For this analysis, there are two continuous numerical variables. The response variable of residence time is not normally distributed. The normality of the explanatory variable is determined with the following code:

hist(Migration$TotalLength)



library(car)
qqPlot(Migration$TotalLength)



shapiro.test(Migration$TotalLength)

##
## Shapiro-Wilk normality test
##
## data: Migration$TotalLength
## W = 0.99429, p-value = 0.8332

Total length is normally distributed. Before conducting the analysis, the relationship between residence time and total length can be viewed with a scatter plot.

ggplot(Migration, aes(TotalLength, Residence.Time)) +
 geom\_smooth(method = "lm", se = TRUE, col = "black") +
 geom\_point(size = 3, col = "firebrick") +
 labs(x = "Total Length (mm)", y = "Residence Time (Days)") +
 theme\_classic()



*Question:* Based upon the scatter plot, does there appear to be a relationship between residence time and total length of sicklefin redhorse?

*Answer:*

A linear regression analysis can be conducted. The null hypothesis is that there is no significant relationship between residence time and total length.

attach(Migration)
MigrationRegression <- lm(Residence.Time ~ TotalLength)
summary(MigrationRegression)

##
## Call:
## lm(formula = Residence.Time ~ TotalLength)
##
## Residuals:
## Min 1Q Median 3Q Max
## -19.402 -12.343 -4.822 9.394 56.775
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.88744 22.76155 1.357 0.177
## TotalLength -0.02101 0.04578 -0.459 0.647
##
## Residual standard error: 15.34 on 145 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared: 0.001451, Adjusted R-squared: -0.005436
## F-statistic: 0.2107 on 1 and 145 DF, p-value: 0.6469

*Question:* Based upon the results of the analysis and the scatter plot, what conclusion is drawn.

*Answer:*

*Question:* Summarize all of the findings from the activity thus far.

*Answer:*

### Hypothesis: There is no relationship between residence time and fish size when accounting for differences between males and females.

Let’s conduct a more thorough investigation of the data. Because residence time differed among males and females, this may be a confounding variable in the analysis of residence time and total length. The relationship between residence time and total length may differ among males and females. To do this, the data set is separated by males and females using the following code:

MaleMigration <- subset(Migration, Sex=="male")
FemaleMigration <- subset(Migration, Sex=="female")

The relationship between residence time and total length will now be analyzed separately for males and females. A regression analysis for males and a scatter plot of the relationship is conducted with the following R code:

attach(MaleMigration)
MaleRegression <- lm(Residence.Time ~ TotalLength)
summary(MaleRegression)

##
## Call:
## lm(formula = Residence.Time ~ TotalLength)
##
## Residuals:
## Min 1Q Median 3Q Max
## -23.467 -10.354 -2.096 8.031 39.728
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -66.91854 26.63603 -2.512 0.013817 \*
## TotalLength 0.19497 0.05447 3.579 0.000564 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.61 on 88 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared: 0.1271, Adjusted R-squared: 0.1172
## F-statistic: 12.81 on 1 and 88 DF, p-value: 0.0005636

ggplot(MaleMigration, aes(TotalLength, Residence.Time)) +
 geom\_smooth(method = "lm", se = TRUE, col = "black") +
 geom\_point(size = 3, col = "firebrick") +
 labs(x = "Total Length (mm)", y = "Residence Time (Days)") +
 theme\_classic()



*Question:* What conclusion can be drawn from the results of the analysis?

*Answer:*

The relationship between residence time and total length for females is investigated with a regression analysis, and a scatter plot of the relationship is created with the following R code:

attach(FemaleMigration)
femaleRegression <- lm(Residence.Time ~ TotalLength)
summary(femaleRegression)

##
## Call:
## lm(formula = Residence.Time ~ TotalLength)
##
## Residuals:
## Min 1Q Median 3Q Max
## -7.0119 -3.4003 -0.8584 2.8616 12.0784
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.505535 13.301087 0.264 0.793
## TotalLength 0.009031 0.026084 0.346 0.731
##
## Residual standard error: 4.833 on 55 degrees of freedom
## Multiple R-squared: 0.002175, Adjusted R-squared: -0.01597
## F-statistic: 0.1199 on 1 and 55 DF, p-value: 0.7305

ggplot(FemaleMigration, aes(TotalLength, Residence.Time)) +
 geom\_smooth(method = "lm", se = TRUE, col = "black") +
 geom\_point(size = 3, col = "firebrick") +
 labs(x = "Total Length (mm)", y = "Residence Time (Days)") +
 theme\_classic()



*Question:* What conclusion can be drawn from the results of the analysis?

*Answer:*

The most effective way to visualize the differences between males and females in the relationship between residence time and total length is plot both relationships on the same graph. This is done with the following R code:

attach(Migration)
cols <- c("pink", "blue")
ggplot(Migration, aes(x = TotalLength, y = Residence.Time, color = Sex)) +
 geom\_point() +
 geom\_smooth(method = "lm", fill = NA)+ scale\_color\_manual(values = cols)+
 labs(x="Total Length (mm)", y="Residence Time (Days)")+
 theme\_classic()



*Question:* Summarize the relationship between residence time and total length when accounting for the effect of sex on the relationship. Speculate as to why this relationship exists.

*Answer:*

### Discussion Questions

Discuss the following questions in small groups. Write the contents of your discussion in the space provided for each question. You are being asked to speculate about these relationships based upon your knowledge from the case study and your general knowledge of ecological concepts from the course.

*Discussion Question:* Speculate as to why residence time and differs among sexes.

*Answer:*

*Discussion Question:* Speculate as to why residence time differs between years.

*Answer:*

*Discussion Question:* Speculate as to why residence time is related to total length for males but not females.

*Answer:*

*Discussion Question:* Evaluate the findings from this lesson in terms of nutrient cycles and the findings of Hudson et al. (2023) from Lesson 2.

*Answer:*