

Lab 9: Cone Handling Handedness

Jackson Anderson

11/11/20

General Information

This lab is due November 11th by 11:59 pm. This lab is worth 10 points. You must upload your .rmd file and knitted file to the dropbox on Canvas. You are welcome and encouraged to talk with classmates and ask for help. However, each student should turn in their own lab assignment and all answers, including all code, needs to be solely your own.

Objective

The goal of this assignment is to give you hands on experience performing general linear model analyses using R. This is not designed to be a quiz of your knowledge of commands in R. I will provide fairly detailed instructions on the code for performing specific analyses, but you will be required to interpret the output from R. Future assignments will contain progressively less detailed instructions on R commands. You will also be asked to write a paragraph of methods and results to give you practice reporting statistical results and to consider how these analyses relate to the data that you will analyse for your independent project.

Biological Background

Believe it or not there are both left and right-handed red squirrels! Right-handed squirrels eat spruce cones by holding the cone primarily in their left hand but all the spinning of the cone with their right hand. Left-handed squirrels do all the spinning with their left hand. Even more intriguing is the fact that there are both dextral and sinistral spruce cones! That is, some spruce cones have scales that spiral in one direction, while other cones spiral in the other direction. This led me to the hypothesis that there might be some sort of coevolution between squirrel handedness and chirality (the direction of spiraling) in spruce cones. One of the first tests of the hypothesis would be to see whether right-handed squirrels eat sinistral cones faster than dextral cones whereas left-handed squirrels eat dextral cones faster than sinistral cones. It takes a long time to collect these sorts of data with squirrels and it is hard to control for the confounding effect of the size of the cone so I conducted some preliminary experiments using some simulated squirrels (field assistants). The assistants did not actually eat the cones but they were instructed to husk the cones as quickly as possible using either a left- or right-handed technique similar to squirrels. The time taken to husk the entire cone was recorded and I provided a competitive atmosphere to try and give them extra motivation. The length (mm) and width (mm) of each cone was measured to account for variation in husking time that was due to the size of the cone. I also recorded the trial number for each “squirrel” in case they got better with more experience. The data from this experiment are included in the comma separated file “ConeHandling2.csv” on Canvas.

General Instructions

In this assignment you will be required to do three things. First, follow the detailed instructions to perform the following analyses using R using this R Markdown file. Be sure to carefully annotate/comment this file with your own personal notes explaining what was done and what the output means as well as answering the questions contained within the specific instructions. At the end of this document, write one paragraph of methods and one paragraph of results describing the analyses that were performed. The methods should refer only to statistical methods. The results should be brief and should include all necessary statistics (refer to handout on 'Presenting Results' on StatsTree <https://static1.squarespace.com/static/5eb33c095018927ea433a883/t/5f5798bf35cfcb0f5c373ab3/1599576336508/Presenting+Results.pdf>). Please pay attention to style and try and roughly follow the format of a scientific article.

Specific Instructions:

Open R and change your working directory to whatever folder you have been using for this class using either the menu or using the `setwd()` command.

Importing the Data and Basic Descriptive Statistics

Download the ConeHandling.csv file from Canvas and import it into your workspace.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1
```

```
## v ggplot2 3.3.2      v purrr   0.3.4
## v tibble  3.0.3      v dplyr  1.0.1
## v tidyr   1.1.1      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflic
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
coneDat <- read_csv("~/Repos/School/BioStats/Lab9/ConeHandling.csv")
```

```
## Parsed with column specification:
```

```
## cols(
##   Squirrel = col_character(),
##   Trial = col_double(),
##   Length = col_double(),
##   Width = col_double(),
##   Handed = col_character(),
##   Chirality = col_character(),
##   Time = col_double()
## )
```

Check to make sure that datafile was imported correctly. There should be 7 variables

```
length(coneDat)
```

```
## [1] 7
```

...that are named (Squirrel, Trial, Length, Width, Handed, Chirality and Time)

```
names (coneDat)
```

```
## [1] "Squirrel" "Trial" "Length" "Width" "Handed" "Chirality"
## [7] "Time"
```

Also check that there are 39 rows of data.

```
length(coneDat$Trial)
```

```
## [1] 39
```

Summarize the data using:

```
summary (coneDat)
```

```
##      Squirrel      Trial      Length      Width
## Length:39      Min.   :1.000      Min.   :35.00      Min.   :10.00
## Class :character 1st Qu.:2.000      1st Qu.:41.95      1st Qu.:12.85
## Mode  :character Median :3.000      Median :45.30      Median :14.30
##              Mean  :2.769      Mean  :45.32      Mean  :13.86
##              3rd Qu.:4.000      3rd Qu.:49.45      3rd Qu.:14.80
##              Max.   :6.000      Max.   :57.00      Max.   :16.30
##      Handed      Chirality      Time
## Length:39      Length:39      Min.   : 61.0
## Class :character Class :character 1st Qu.: 82.5
## Mode  :character Mode  :character Median :105.0
##              Mean  :115.3
##              3rd Qu.:146.0
##              Max.   :209.0
```

We will convert Squirrel, Handed and Chirality into factors.

```
coneDat$Squirrel<-as.factor(coneDat$Squirrel)
coneDat$Handed<-as.factor(coneDat$Handed)
coneDat$Chirality<-as.factor(coneDat$Chirality)
```

```
summary(coneDat)
```

```
##      Squirrel      Trial      Length      Width      Handed
## Mireille : 6      Min.   :1.000      Min.   :35.00      Min.   :10.00      L:20
## Mark      : 5      1st Qu.:2.000      1st Qu.:41.95      1st Qu.:12.85      R:19
## Adi       : 4      Median :3.000      Median :45.30      Median :14.30
## Christina: 4      Mean    :2.769      Mean    :45.32      Mean    :13.86
## Mike      : 4      3rd Qu.:4.000      3rd Qu.:49.45      3rd Qu.:14.80
## Moe       : 4      Max.    :6.000      Max.    :57.00      Max.    :16.30
## (Other)   :12
## Chirality      Time
## D:19      Min.   : 61.0
## S:20      1st Qu.: 82.5
##              Median :105.0
##              Mean    :115.3
##              3rd Qu.:146.0
##              Max.    :209.0
##
```

Notice the names of my “squirrels”. The codes for Handed correspond to cones that were husked in a left-handed (L) or right-handed (R) way. Chirality refers to whether the bracts on the cone were configured in a dextral (D) or sinistral (S) direction. Recall that Time refers to the length of time taken to completely

husk the cone and is measured in seconds. The Length and Width of the cone are measured in mm. Width seems to work better here as a measure of size so we will use that from here on. Trial also was not a very good predictor of cone handling time suggesting that my squirrels did not improve with practice! We will not consider it further.

What is the mean width of cones? Calculate it. And calculate the se for the width of cones as

```
meanConeWidth <- mean(coneDat$Width)
seWidth <- sd(coneDat$Width)/sqrt(length(coneDat$Width))
```

What was the average length of time needed to “eat” a cone? As practice, calculate the se for the length of time needed to eat a cone

```
meanTime <- mean(coneDat$Time)
seTime <- sd(coneDat$Time)/sqrt(length(coneDat$Time))
seTime
```

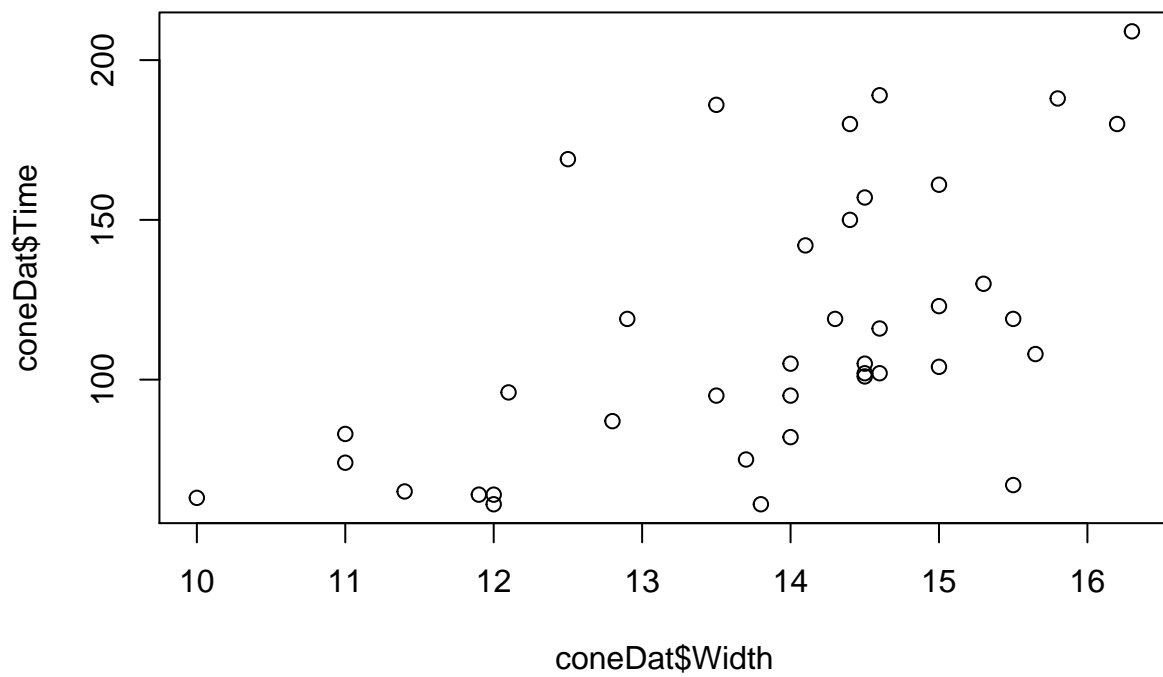
```
## [1] 6.823488
```

Include these basic descriptive statistics at the start of your results section, but remember to focus on biology and not statistics.

Exploratory Graphical Analysis

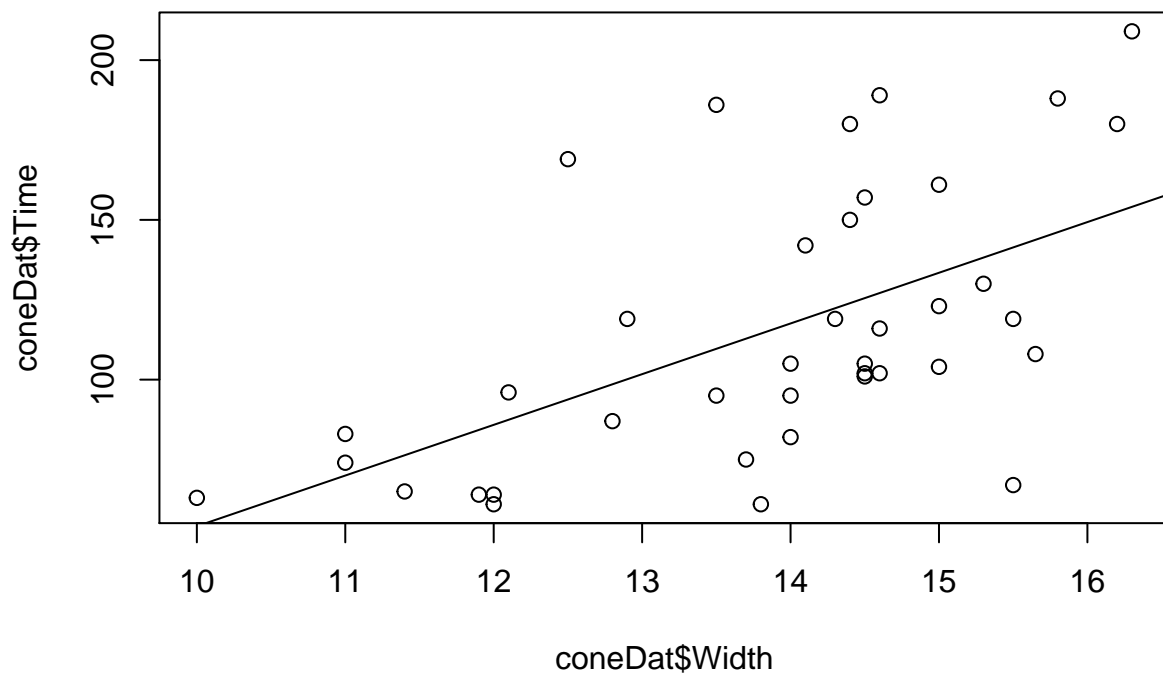
We suspect that there might be a relationship between the width of the cone and the length of time taken to eat it. Explore this graphically first...

```
plot(coneDat$Width, coneDat$Time)
```



Does the plot look linear? Try adding a line

```
plot(coneDat$Width, coneDat$Time, abline(lm(Time~Width, data = coneDat)))
```



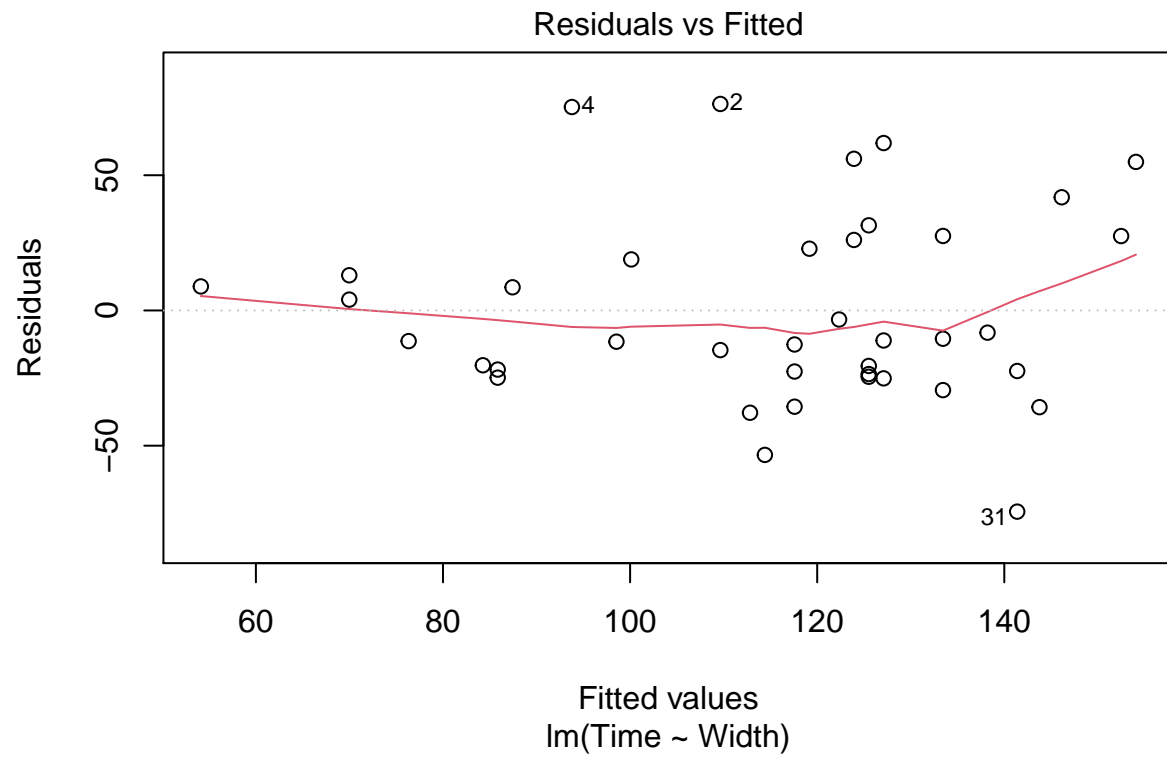
It looks like the line underestimates Time for large cones. We will make a simple model and check the diagnostics for this relationship.

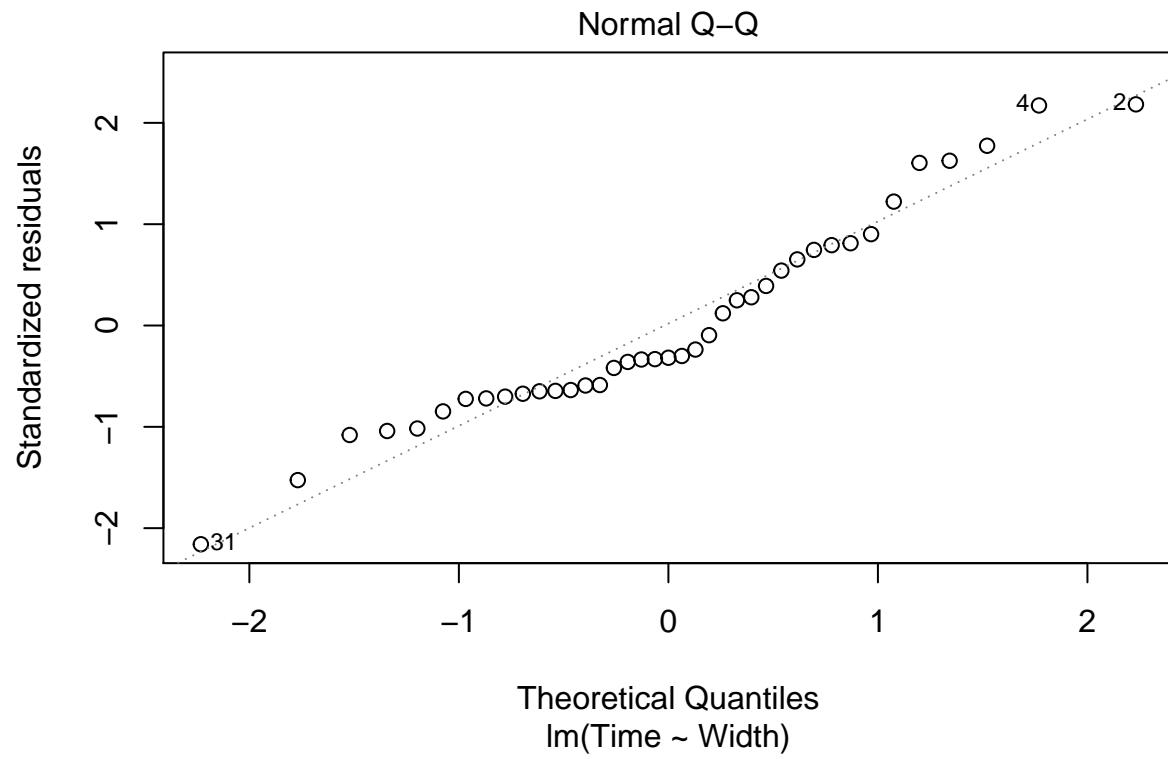
```
model<-lm(Time~Width, data = coneDat)
```

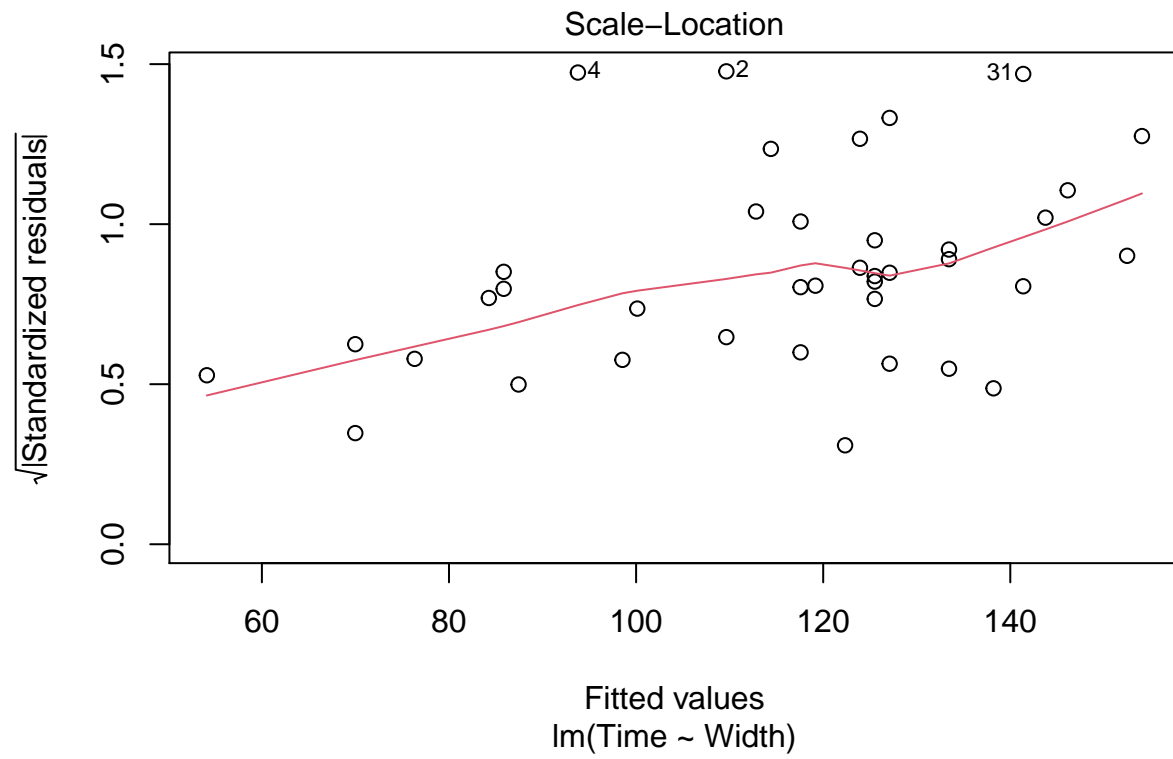
Regression Diagnostics

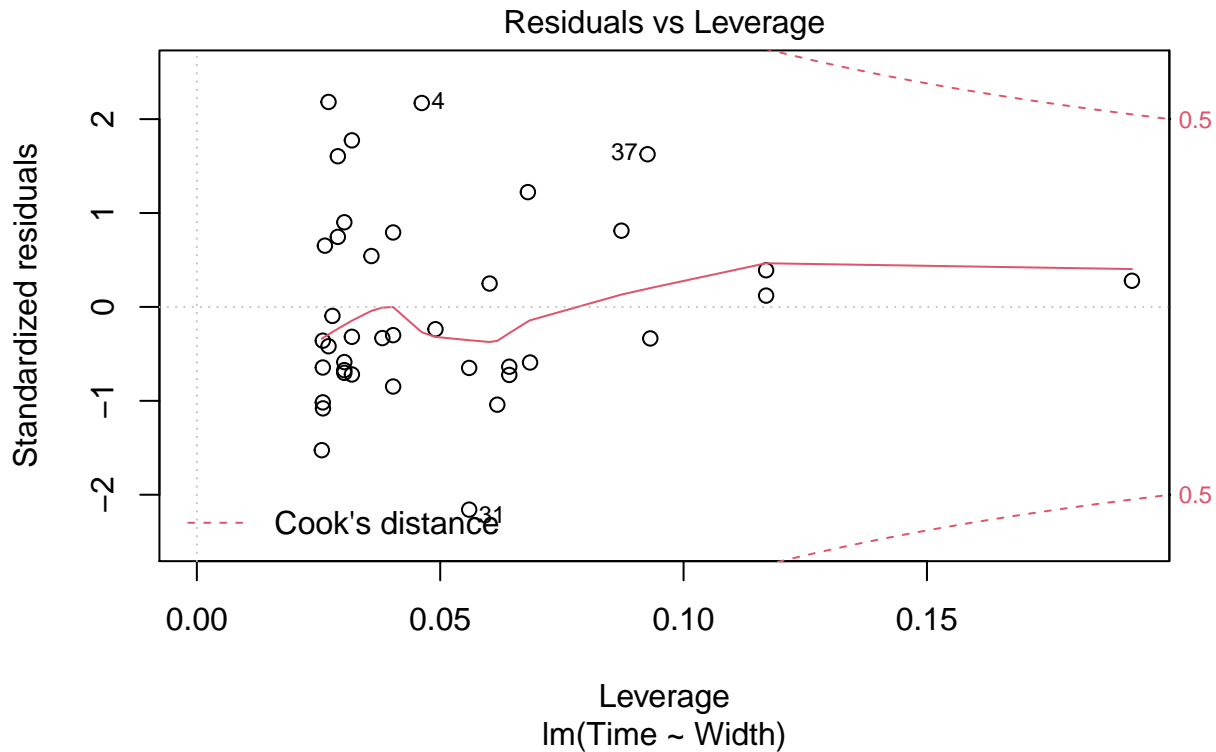
Check the diagnostics using

```
plot(model)
```







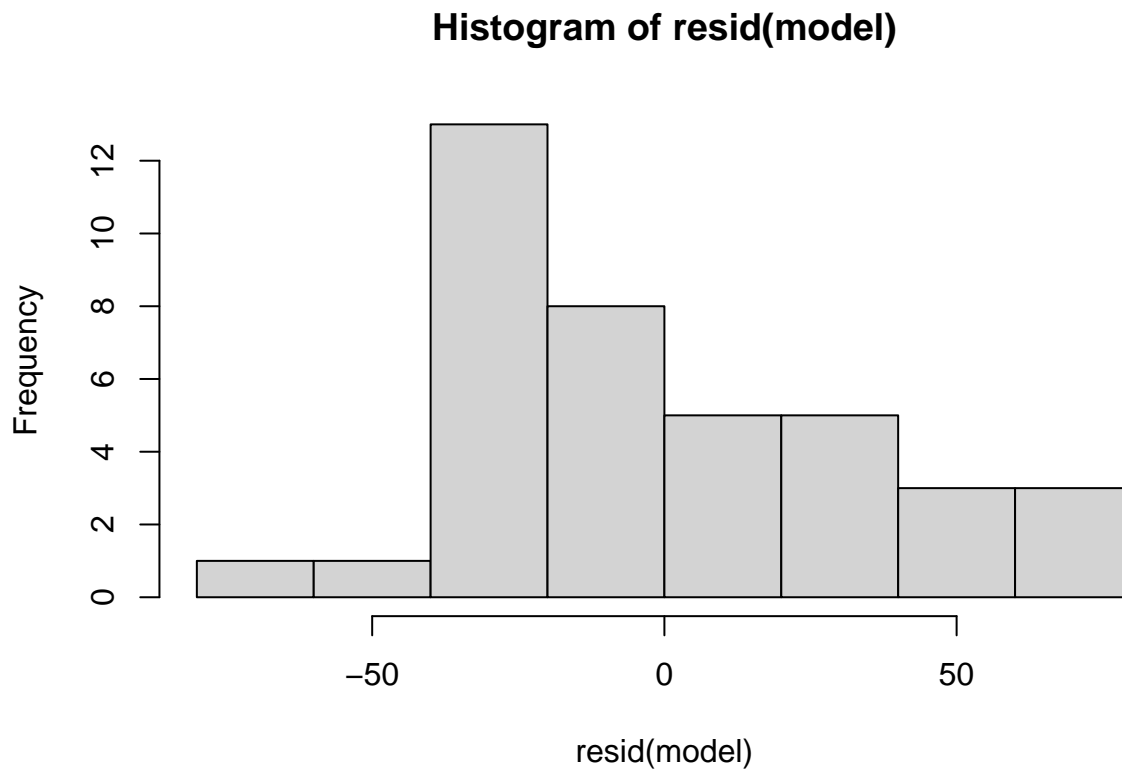


For practice, comment on how these diagnostic plots look. Does the error in the model increase with fitted values? Do the data look normal in the qq plot?}}

There seems to be a slight increase in residual values with an increase in fitted values; overall, though, the distribution of residuals around 0 does not really bother me. There does not seem to be too much deviation of resid values for the qqplot; however, as resid values get more extreme, they do tend to skew a little, specially in the positive direction

Supplement this analysis with a histogram of the residuals

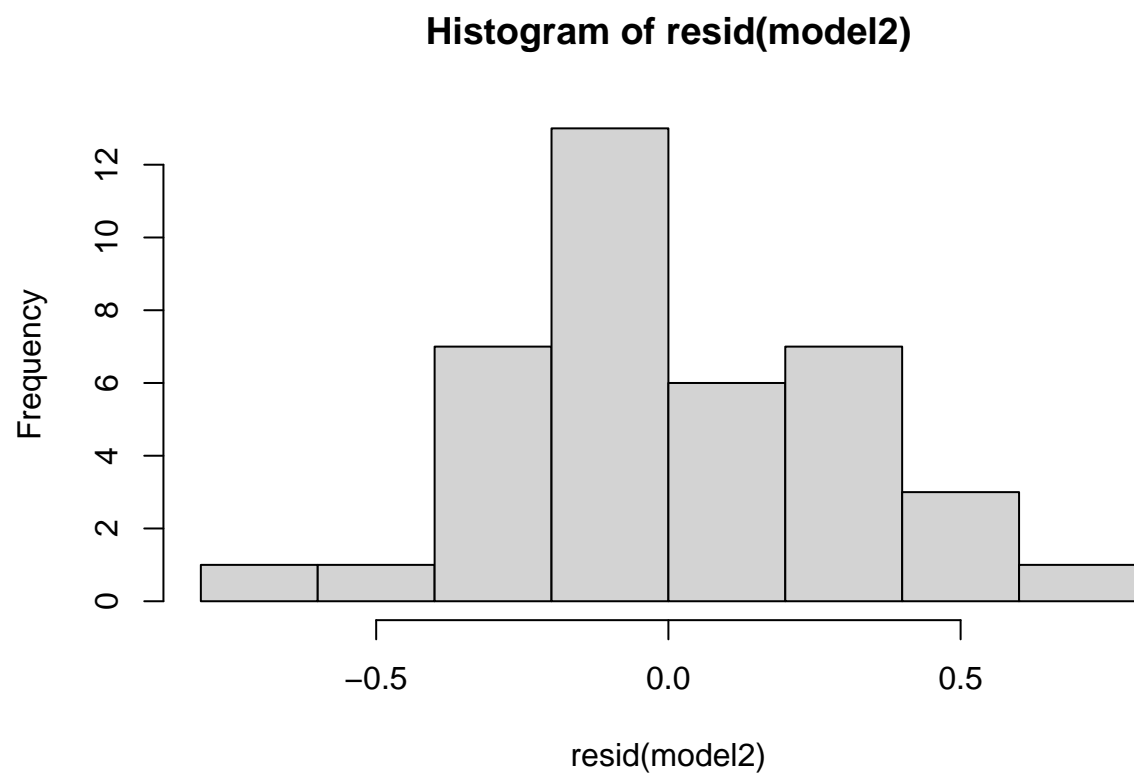
```
hist(resid(model))
```



Do the residuals look normal? *There is a skew towards the positive end of the residual spectrum*

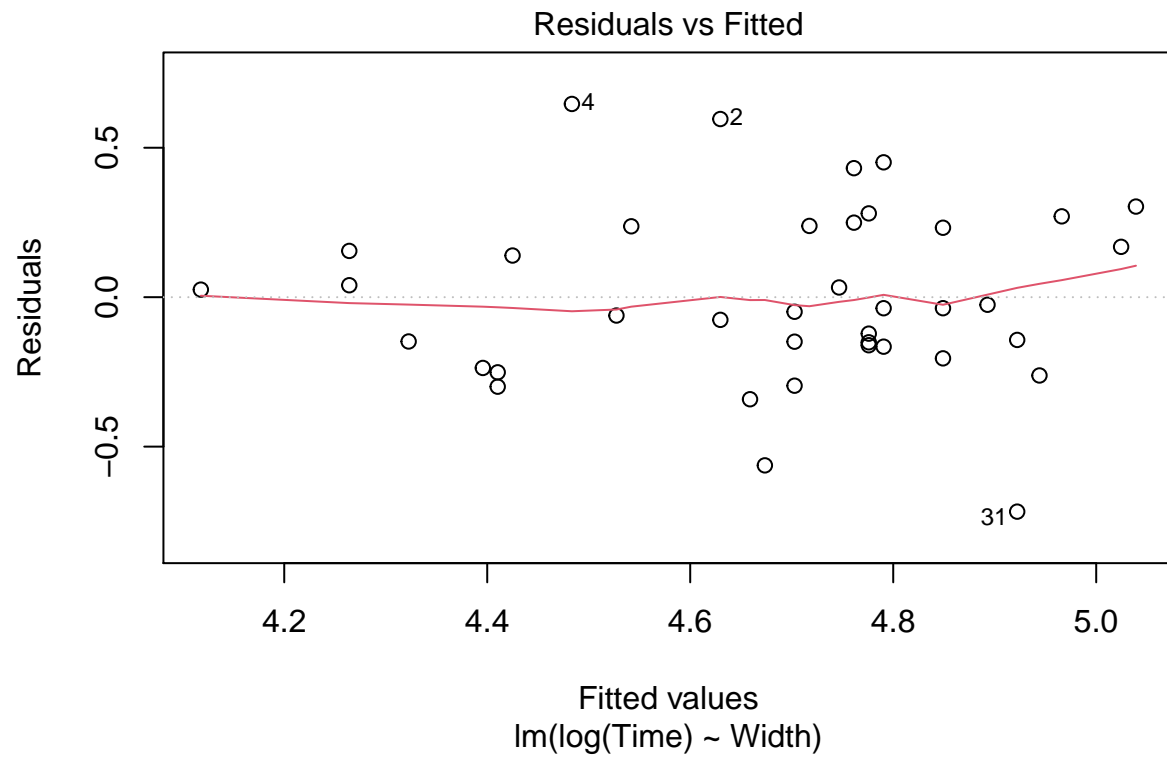
The positive skew suggests that a log transformation might be a good idea.

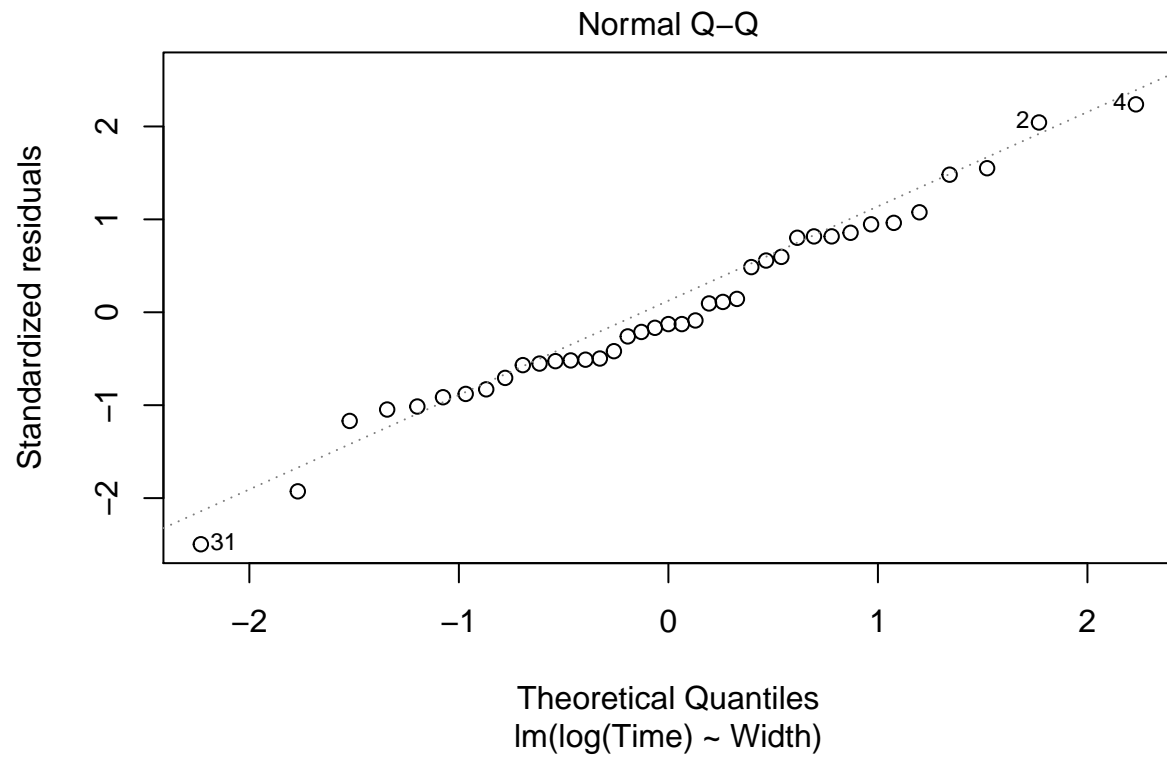
```
model2<-lm(log(Time)~Width, data = coneDat)
hist(resid(model2))
```

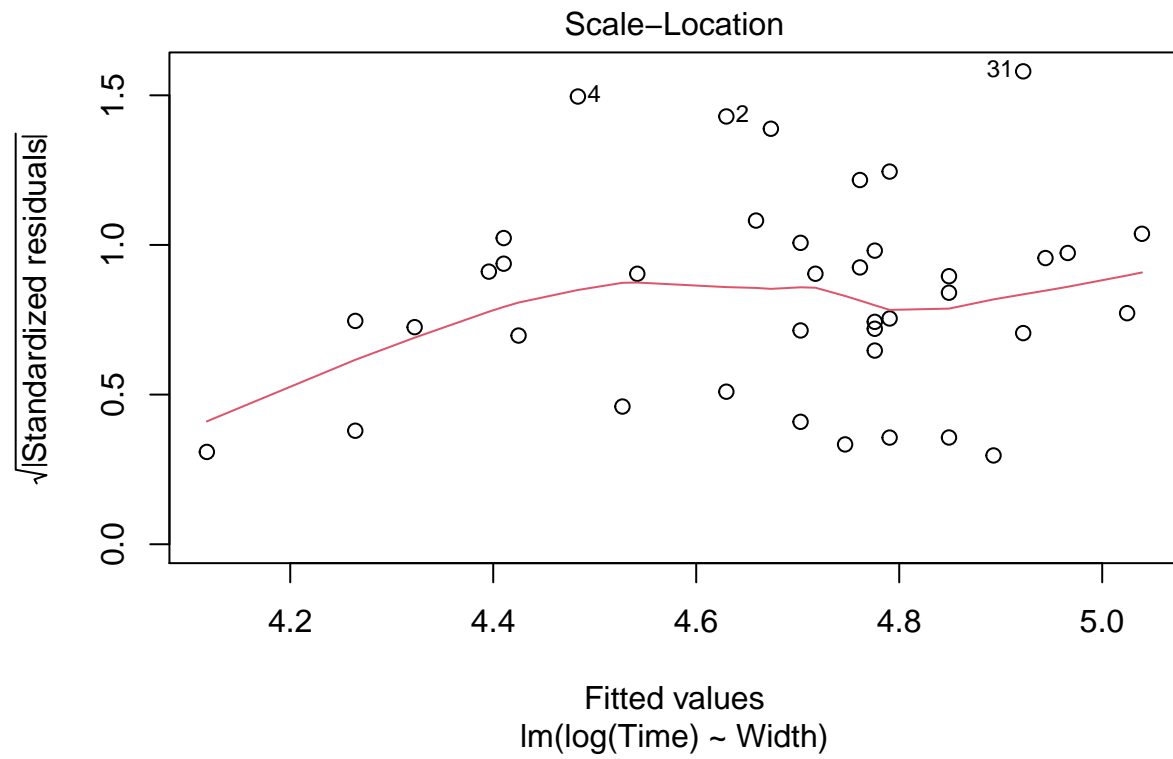


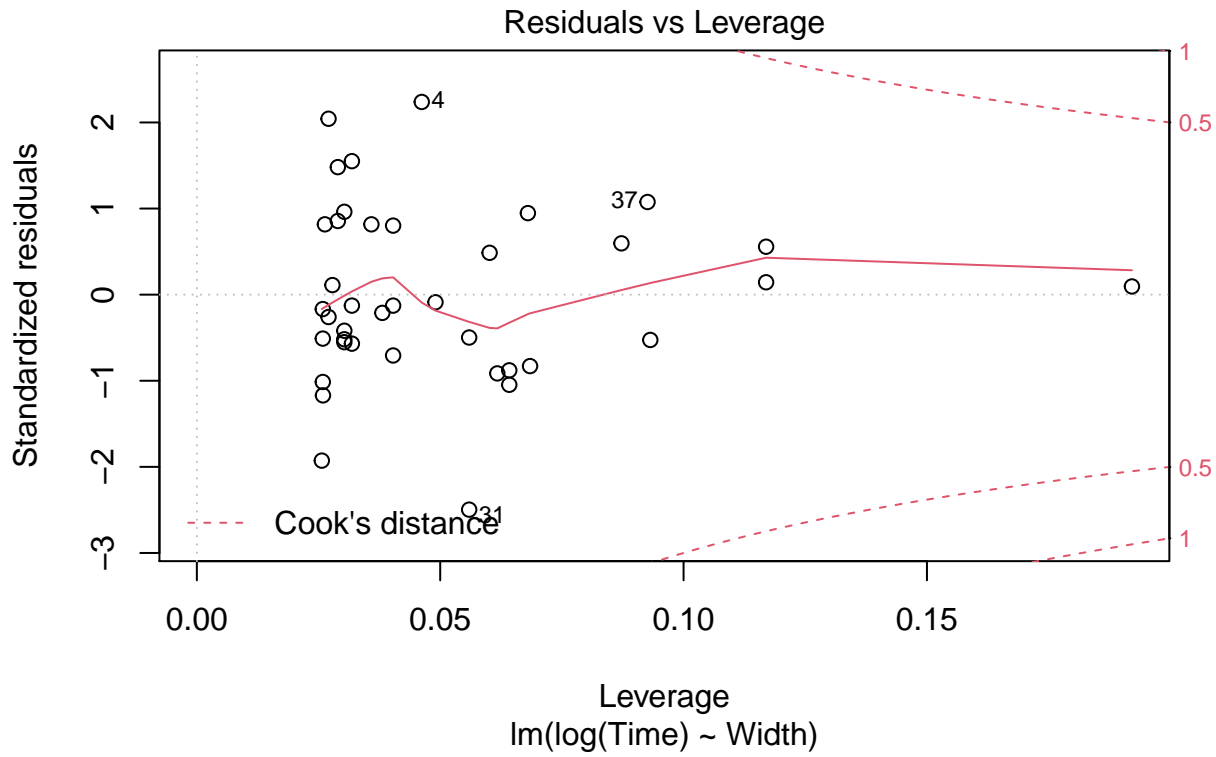
Assess the rest of the diagnostics.

```
plot(model2)
```









These residuals look a bit better. We'll come back to this at the end. Include this transformation in your methods and don't forget that we are dealing with $\log(\text{Time})$ from now on.

glm Results

We can now look at the results of the model

```
summary(model2)
```

```
##
## Call:
## lm(formula = log(Time) ~ Width, data = coneDat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71748 -0.16319 -0.03696  0.23477  0.64645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.6555     0.4363   6.086 4.80e-07 ***
## Width         0.1462     0.0313   4.672 3.87e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2958 on 37 degrees of freedom
## Multiple R-squared:  0.371, Adjusted R-squared:  0.354
```



```
## F-statistic: 21.82 on 1 and 37 DF,  p-value: 3.873e-05
```

Question 1 (1 point) Is Width a significant predictor of handling time? Include these results in your results write-up (below). Was it a positive or a negative relationship?

Width had a significant effect on time, in this model ($p < .001$); the relationship was positive

Question 2 (1 point) What is the regression equation for $\log(\text{Time})$ (i.e. $\log(\text{Time}) = ?? + ?? \text{ Width}$)?

The regression equation is: $\log(\text{time}) = 2.655 + 0.1462\text{Width}$

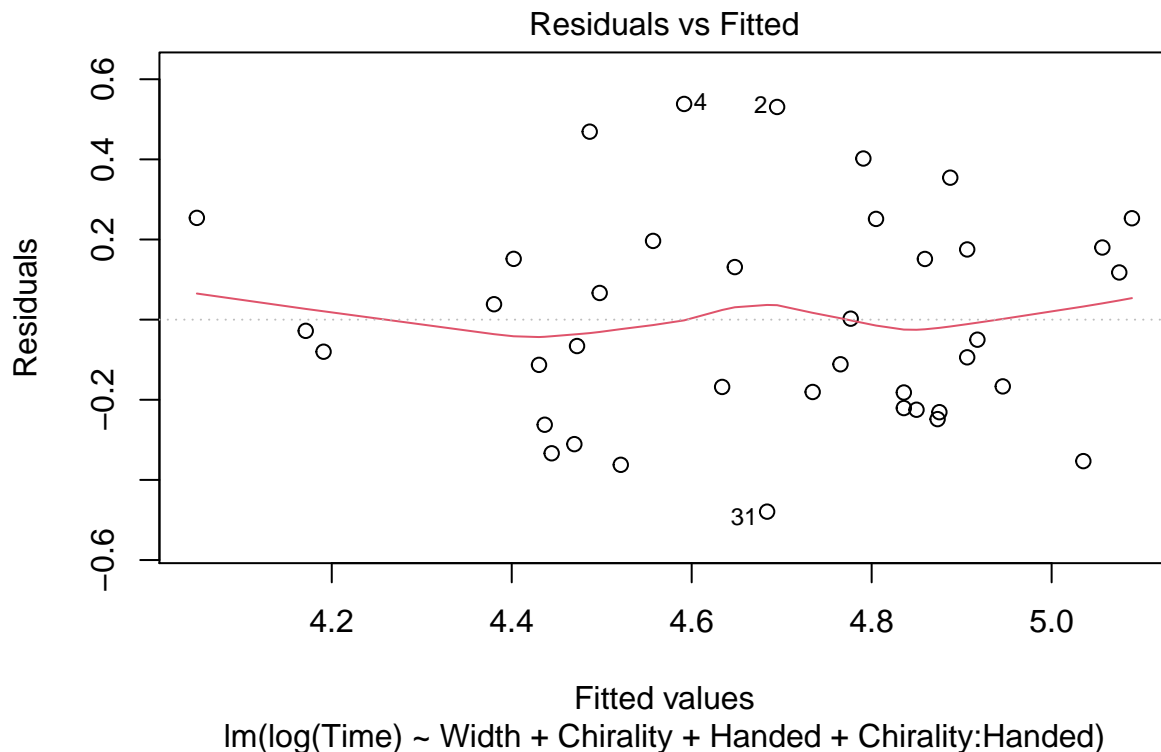
Interaction Model

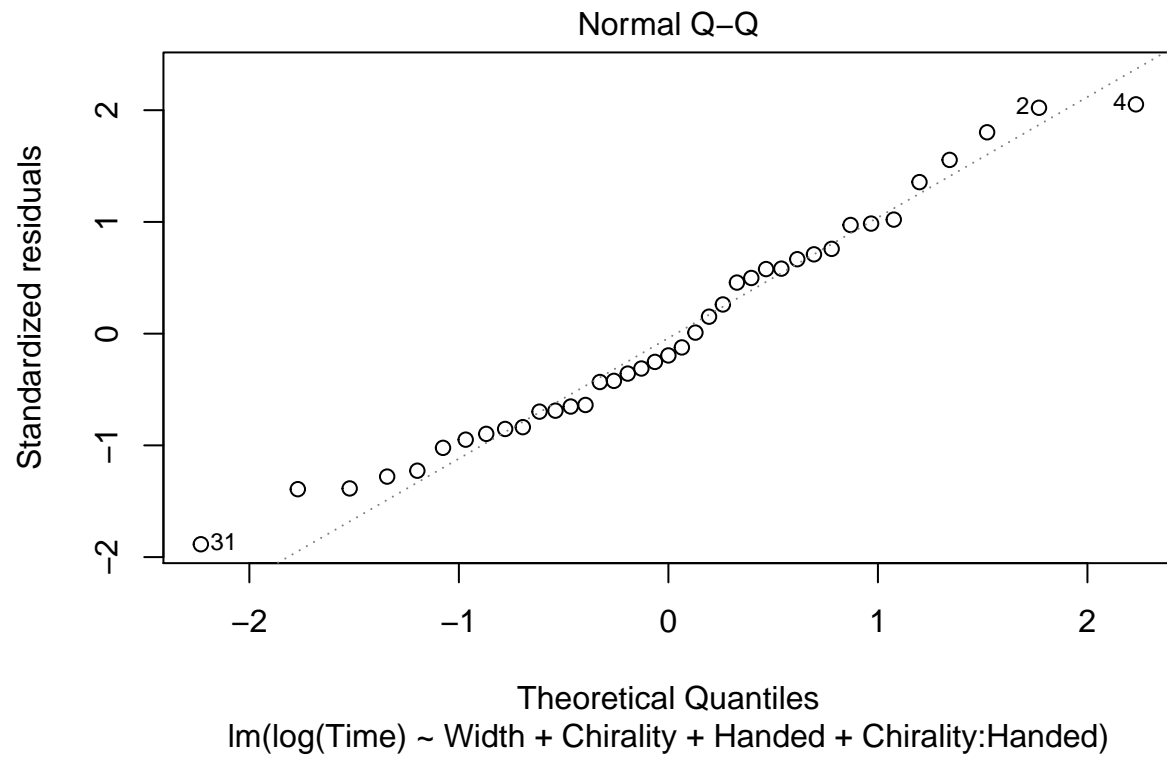
So it will be important to consider the confounding effect of Width when testing for effects of Chirality and Handedness. So, we are now ready to test for the effects that we are really interested in. Remember that our hypothesis is that the time taken to eat a cone does not just depend on the chirality of the cone, but that this “chirality effect” depends on whether the squirrel is left- or right-handed. That is, we are interested in testing for an interaction between handedness and chirality. We also want to control for effects of the size of the cone (Width). We will therefore define the following linear model.

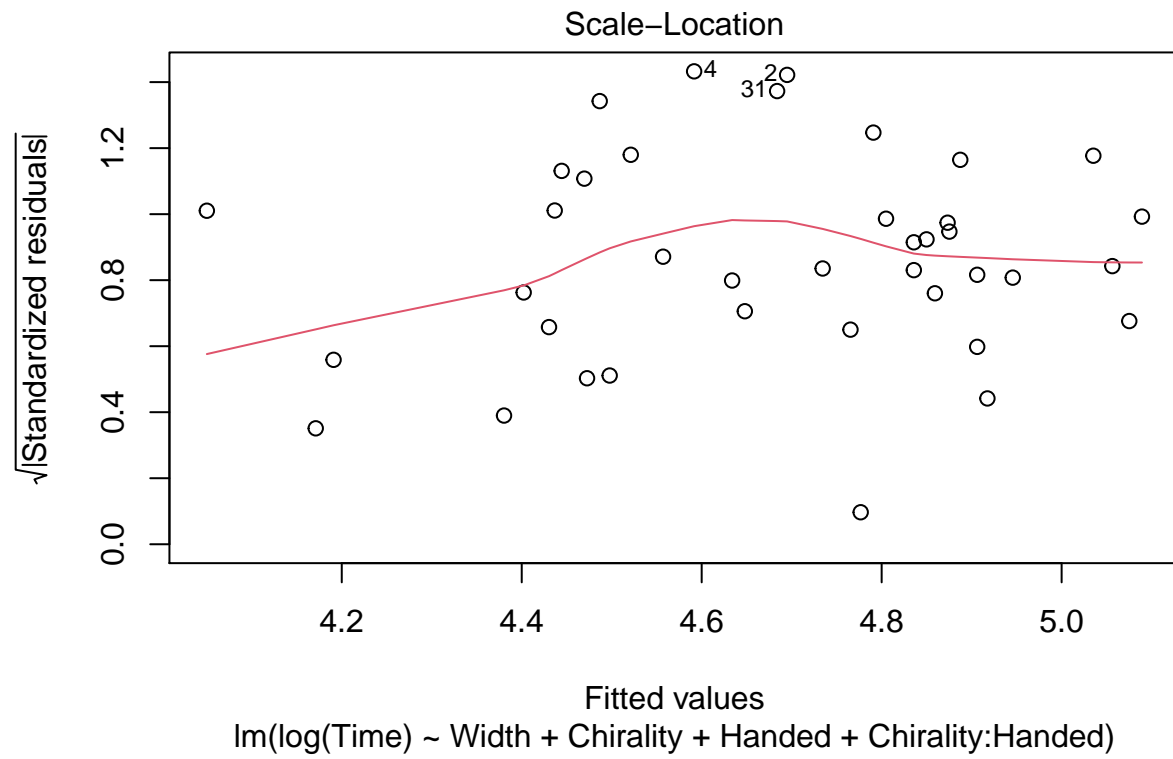
```
model3<-lm( log(Time) ~ Width + Chirality + Handed + Chirality:Handed,  
            data=coneDat,  
            na.action=na.omit)
```

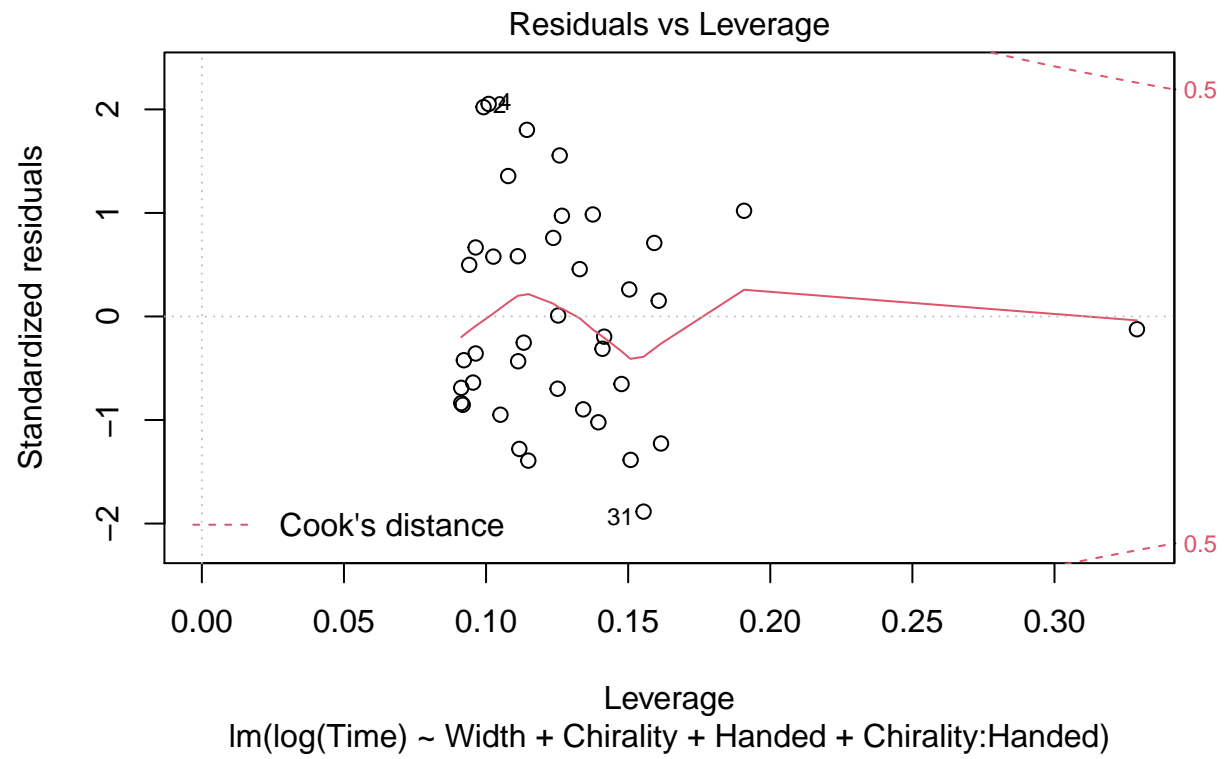
Check the diagnostics for this new model as we did above for model and model2, including a histogram of the residuals.

```
#Diagnostics  
plot(model3)
```

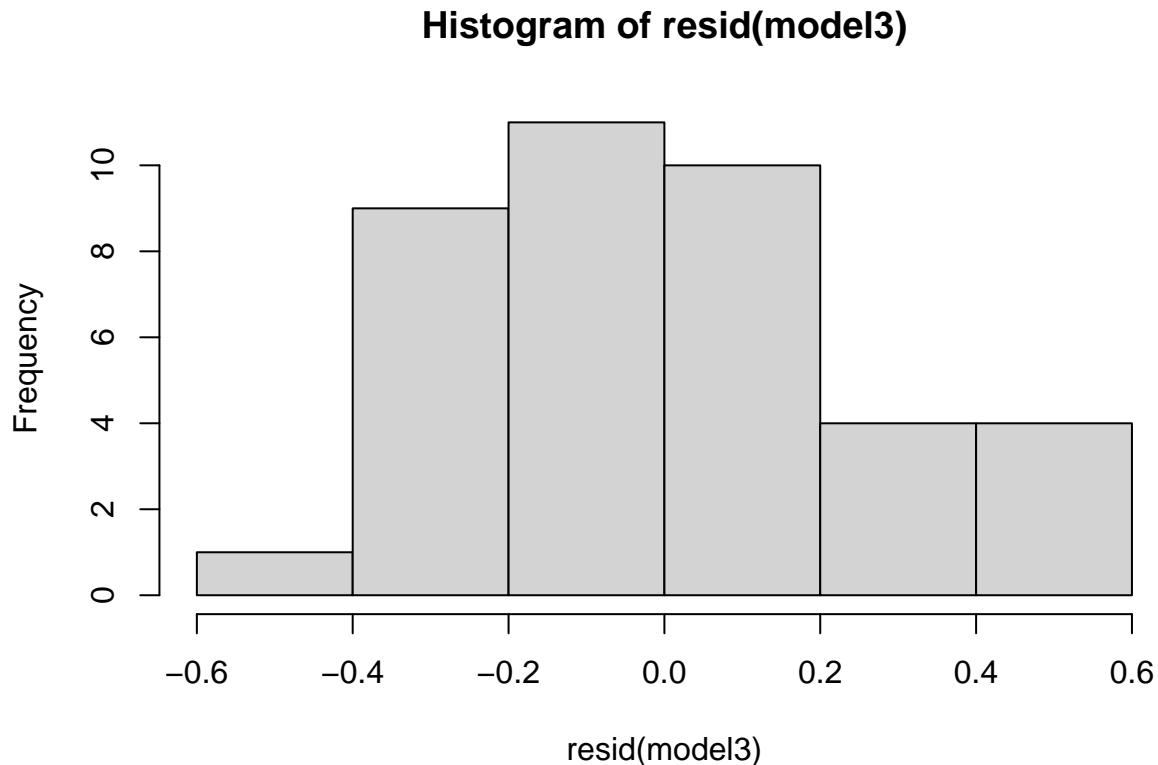








```
#histogram  
hist(resid(model13))
```



Question 3 (1 point) Do the residuals look normal? Are there signs of heteroscedasticity? Do any observations stand out as potential outliers?

Residual distribution looks relatively normal; not concerning. There is a slight skew towards positive residual values, but nothing that warrants concern. Observations 2, 4, and 31 are potential outliers

The absolute values of these Cook's distances are not large so we will proceed. You do not need to include the details of these diagnostics in your methods section but should state which assumptions were examined.

You can now proceed to assess the significance of the model.

```
summary (model3)
```

```
##
## Call:
## lm(formula = log(Time) ~ Width + Chirality + Handed + Chirality:Handed,
##     data = coneDat, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4792 -0.2012 -0.0500  0.1776  0.5383
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.83107    0.41474   6.826 7.44e-08 ***
## Width          0.14084    0.03029   4.650 4.86e-05 ***
## ChiralityS     -0.33028    0.12442  -2.654  0.0120 *
## HandedR        -0.06841    0.13032  -0.525  0.6030
```

```
## ChiralityS:HandedR  0.36118    0.17882    2.020    0.0513 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2766 on 34 degrees of freedom
## Multiple R-squared:  0.4945, Adjusted R-squared:  0.435
## F-statistic: 8.315 on 4 and 34 DF,  p-value: 8.643e-05
```

Question 4 (1 point) Was there a consistent effect of chirality on cone handling time or did this effect depend on handedness of the squirrels? How much of the variation in cone handling time is explained by this model? Which statistics refer to the significance of the model as a whole and which refer just to specific variables in the model?

The effect of chirality on handling time is dependent on the level of handedness in squirrel; the interaction effect was marginally significant ($p = .051$). Approximately 50% of the variation in handling time is explained by this model and its complications. The p value of $8.643e-05$, or the model p value, describes the significance of this entire model, while each variable or interaction effect p -value indicates that variable's effect on Handling time.

Since the interaction is marginally significant we will keep it in the model and will not consider the effects of Chirality and Handedness on their own, henceforth. Write the code for this regression model below, and write out the regression equation for this model on your annotated/commented output as well (i.e. $\log(\text{Time}) = ?? + ??\text{Width} + \dots$). Include the results for this model (not the equation) in your results section. Remember that cone Width was included as a covariate in the model.

#Even after the email from Andrew, still a little confused...it sounds like the instructions want us to

```
coneModel<-lm( log(Time) ~ Width + Chirality + Handed + Chirality:Handed,
              data=coneDat,
              na.action=na.omit)
summary(coneModel)
```

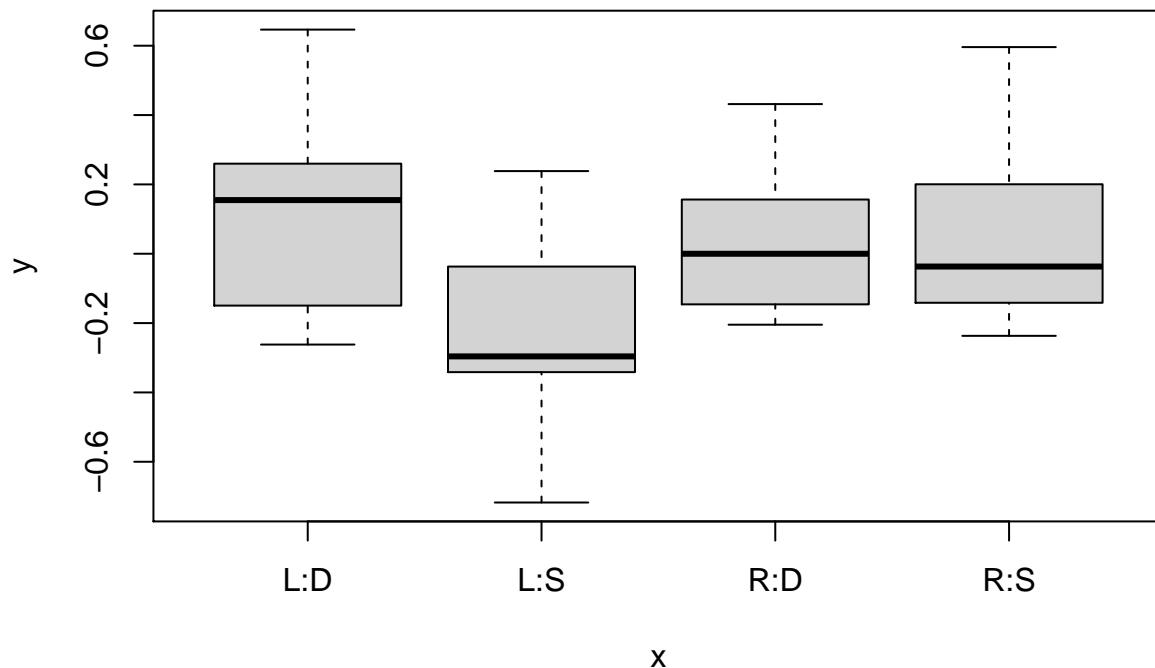
```
##
## Call:
## lm(formula = log(Time) ~ Width + Chirality + Handed + Chirality:Handed,
##     data = coneDat, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4792 -0.2012 -0.0500  0.1776  0.5383
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.83107    0.41474   6.826 7.44e-08 ***
## Width          0.14084    0.03029   4.650 4.86e-05 ***
## ChiralityS     -0.33028    0.12442  -2.654  0.0120 *
## HandedR        -0.06841    0.13032  -0.525  0.6030
## ChiralityS:HandedR 0.36118    0.17882   2.020  0.0513 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2766 on 34 degrees of freedom
## Multiple R-squared:  0.4945, Adjusted R-squared:  0.435
## F-statistic: 8.315 on 4 and 34 DF,  p-value: 8.643e-05
```

The regression equation for this model is: $\log(\text{time}) = 2.831 + .141x\text{Width} - .33x\text{Chirality} - .07x\text{Handed} + .361x\text{Chirality:Handed}$. Overall, this model significantly accounted for the variation seen in the response variable with its several complications. The effect of cone width on handling time was significant, as stated

above ($p < .001$); chirality had a significant effect on handling time ($p < .05$), while handedness did not; the interaction effect of chirality on handedness on handling time was marginally significant ($p = .0513$), and thus the effect of chirality on handling time is dependent on handedness of the squirrel. ## Direction of Effects - Graphics

Finally, remember that we had some specific predictions about how cone handling time should differ depending on the chirality of the cone and the handedness of the squirrel. We predicted that right-handed squirrels should eat sinistral cones faster whereas left-handed squirrels should eat dextral cones faster. The marginally significant interaction indicates that the effect of chirality does depend on whether the squirrel is left or right-handed but we haven't looked to see the direction of the effects. We can do this by using a nifty variation on the plot command

```
plot(coneDat$Handed:coneDat$Chirality, resid(model2))
```



Remember that model2 was our model that examined the effect of Width on Time. So by plotting the residuals of this model with respect to Handedness and Chirality we are looking at the interaction after accounting for the known effects of Width on Time.

Question 5 (1 point) Were the results indicated in the plot as we had suspected? Did Right-handed squirrels (R) eat sinistral cones (S) faster and did left-handed squirrels (L) eat dextral cones (D) faster? Remember that the Y axis here is simply a corrected measure of time (i.e. lower values mean less time).

Based on the plot above, it would seem that our predictions were incorrect with regards to right handed squirrels. However, left handed squirrels ate dextral cones faster than they did sinistral cones, so our predictions proved correct in that case

Include this figure in your results and be sure to discuss the direction of the effects.

You may have noticed that more than one trial was performed per 'squirrel' so that the current analysis is

pseudoreplicated. This is the reason why the residuals do not seem to meet the assumptions of normality and homoscedasticity as well as we might like. In a future assignment we will analyze these same data and correct for this sub-sampling.

Your Methods and Results Sections (3 points) Write one paragraph of methods and one paragraph of results describing the analyses that were performed. The methods should refer only to statistical methods. The results should be brief and should include all necessary statistics (refer to handout on ‘Presenting Results’ on StatsTree <https://static1.squarespace.com/static/5eb33c095018927ea433a883/t/5f5798bf35cfcb0f5c373ab3/1599576336508/Presenting+Results.pdf>). Please pay attention to style and try and roughly follow the format of a scientific article.

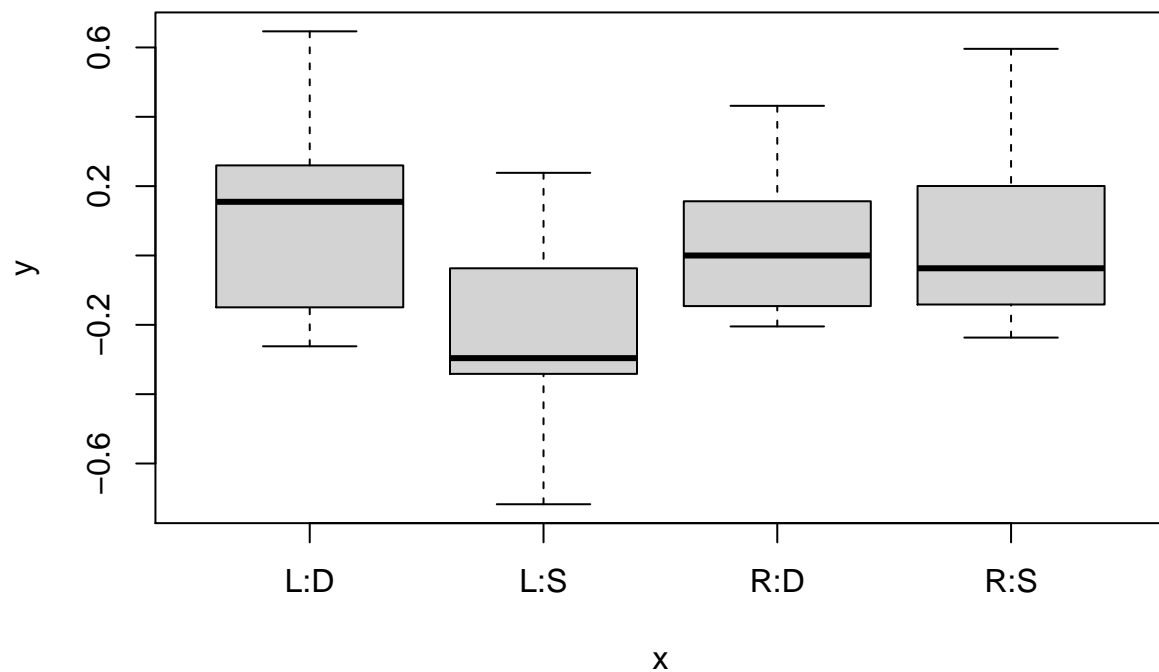
Methods

A generalized linear model (GLM) was used to determine whether squirrel handedness, cone chirality, or a chirality:handedness interaction significantly affected cone husk time; cone width was treated as a covariate within the model framework. The unit of replication was that of the handedness; whether the cone was husked using the right or the left hand. However, it should be noted that each squirrel was responsible for both right- and left-handed husks, and so subsequent analyses were based on these pseudoreplications. Due to slight heteroskedasticity in residual variance within the time variable, it was log transformed. All analyses are based on a log transformed time variable. The GLM and

Results

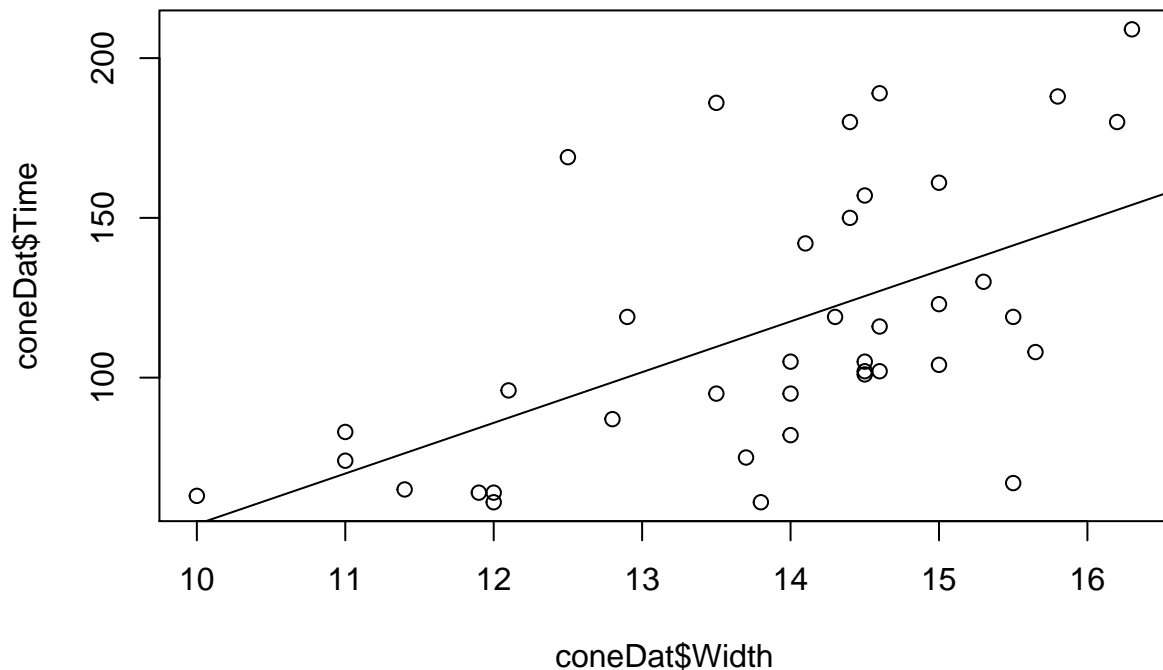
Within our GLM, our predictor values accounted for a significant amount of the variation in cone handling time ($F(4,34) = 8.32, p < 0.001, se = 0.28$). The mean time to husk a cone, regardless of squirrel handedness, was 115 seconds with a standard error of 6.82 seconds. While cone handling time was not significantly affected by cone chirality ($p < .05$) or squirrel handedness ($p < .05$), the interaction effect between the two was significant; left-handed squirrels husked dextral cones faster than they did sinistral cones, while right handed squirrels husked sinistral cones faster than dextral cones ($t(34) = 2.02, p = 0.0513, se = 0.18$) (See Figure Below).

```
plot(coneDat$Handed:coneDat$Chirality, resid(model2))
```

Cone width also proved to be a significant covariate with regards to effects on handling time ($t(34) = 4.65$, $p < 0.001$, $se = 0.03$)

```
plot(coneDat$Width, coneDat$Time, abline(lm(Time~Width, data = coneDat)))
```



Your Data (2 points) Write a couple of sentences that describe the ways in which your own dataset is similar to this one. What is your replicate that would be analogous to a trial? Is this a sub-sample of some larger variable that would be equivalent to the squirrels here? What are your categorical and continuous variables that would be equivalent to Chirality, Handed or cone Width? Are you likely to investigate a similar interaction?

For my data set, the replication level is at that of the total number of species planted or originally counted within each community treatment category; there is no pseudoreplication in this experiment. There are several variables of interest here, but I am most interested in looking at the interaction of plant community composition and fertilizer treatment; community composition (two levels) is the broad treatment effect, while fertilizer treatment is applied to each community. I suspect there will be a significant interaction going on there