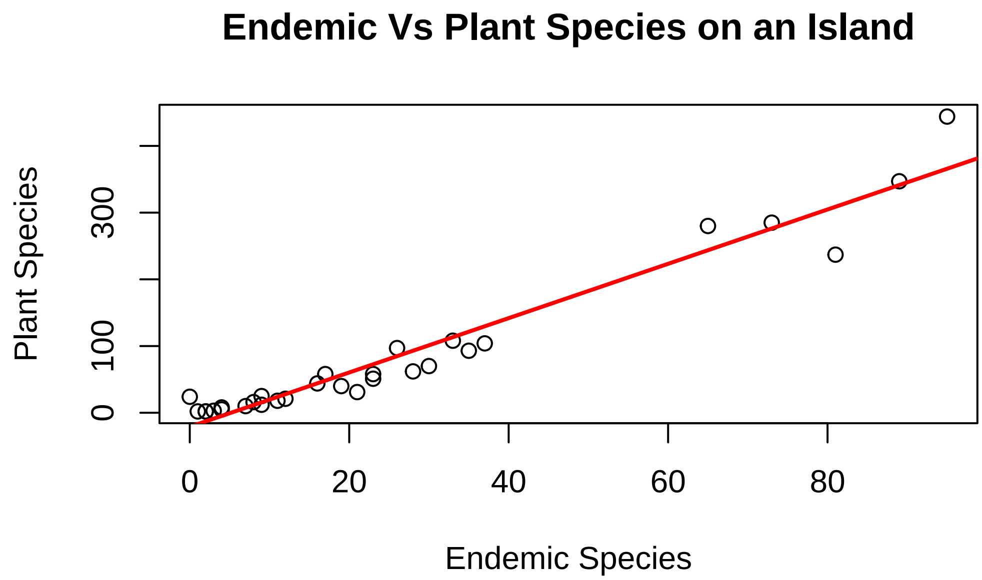
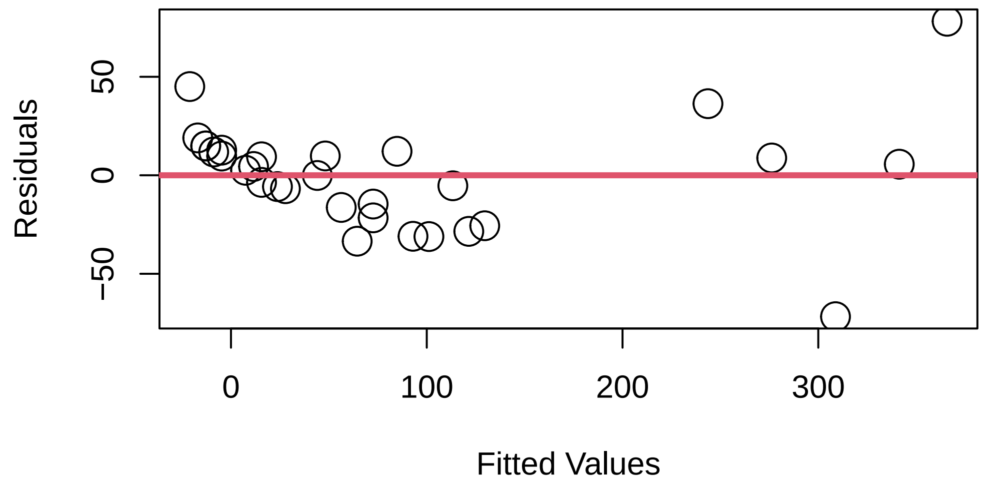
Lab 3

Brad Staples

Question 1






```
    F test to compare two variances

data:  reg1$residuals[group1] and reg1$residuals[!group1]
F = 7.541, num df = 14, denom df = 14, p-value = 0.0005429
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  2.531737 22.461527
sample estimates:
ratio of variances
          7.540999
```

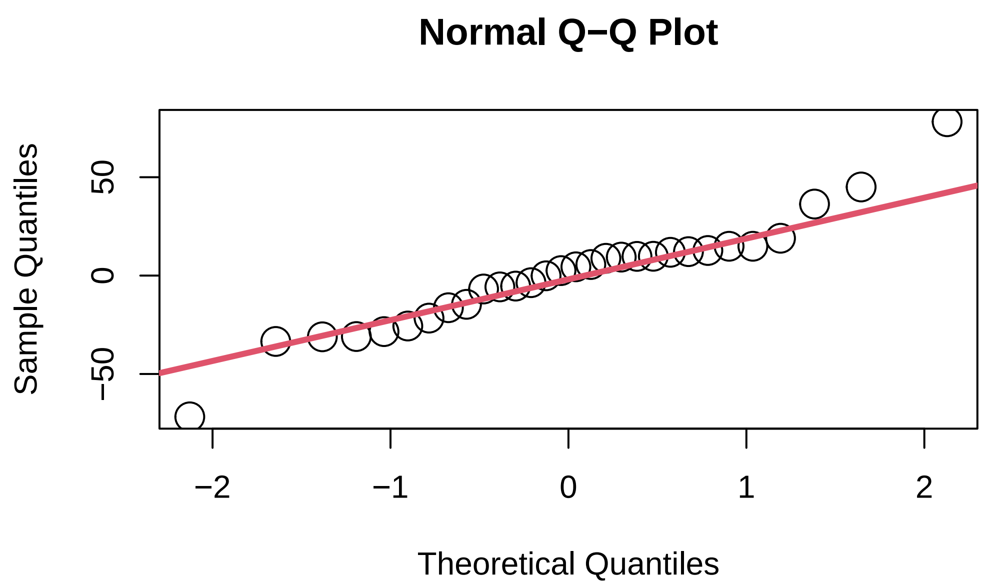
Linearity

From the fitted model scatter-plot we can see that the relationship between endemic species and plant species is not perfectly linear. There is a noticeable dip in the midrange between 20 and 40 on the x-axis, with several higher-end outliers above 60 that the line does not pass through at all, but still pull the regression line upwards. The fitted vs residual graph also shows a slight curve pattern between 0 to 150 on the x-axis. Altogether, these patterns suggest that linearity is not a valid assumption.

Homoskedasticity

The residual plot shows a slight fanning pattern: points on the left are tightly clustered around the center, but the spread increases as fitted values grow. The variance test supports this unequal variance, leading me to conclude that the homoskedasticity assumption is not valid.

```
#qq test, sharpiro.test
qqnorm(reg1$residuals, cex=2)
qqline(reg1$residuals, col=2, lwd=3)
```



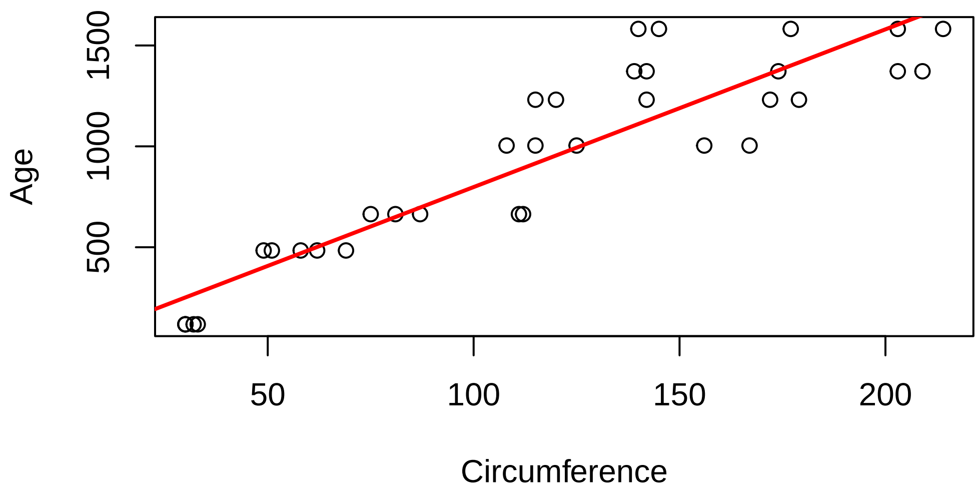

```
shapiro.test(reg1$residuals)
```

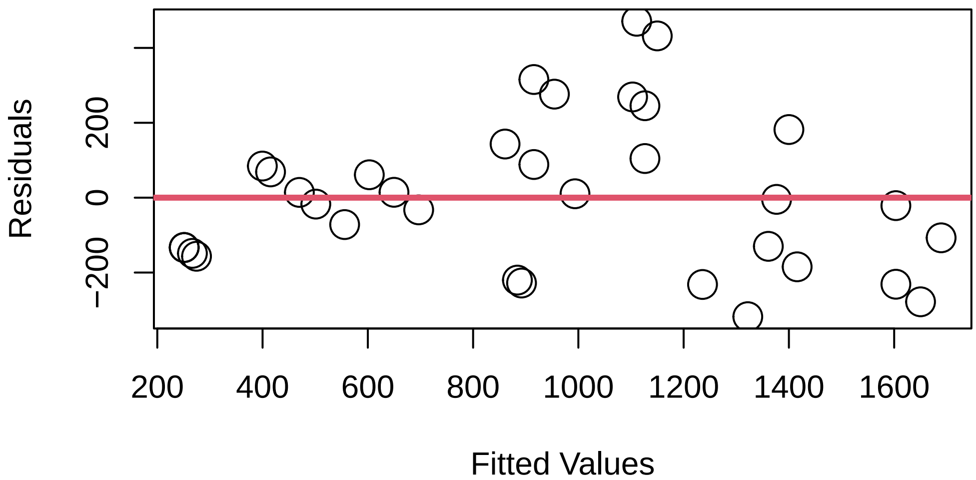
```
	Shapiro-Wilk normality test

data:  reg1$residuals
W = 0.94501, p-value = 0.1241
```

Normality

The QQ plot shows that most points fall close to the line, though there are some deviations at the tails. The line only barely passes through a lot of the middle points and becomes far less accurate on the edge cases. However when we run the Shapiro-Wilk test, we end up with a P-value above 0.05, which means we fail to reject the null hypothesis and the normality assumption is reasonably valid, despite the graph being a little awkward.

Question 2






```
    F test to compare two variances

data:  reg2$residuals[group2] and reg2$residuals[!group2]
F = 3.1363, num df = 16, denom df = 17, p-value = 0.02474
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.162993 8.587241
sample estimates:
ratio of variances
          3.136321
```

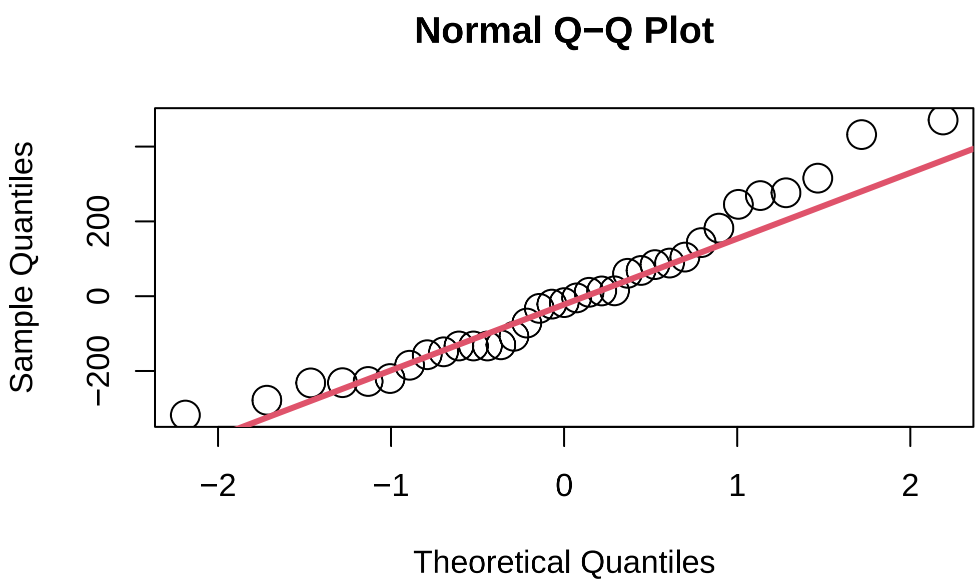
##Linearity

The fitted model shows a lot of the points spread above or below the regression line. Examining the fitted vs residual plot, we can see a distinct n shaped curve with the points in the middle being the highest and the edge cases being the lowest. This pattern suggests a nonlinear relationship between age and circumference and the invalidity of the linearity assumption.

Homoskedasticity

There is a pretty decent fan shape to this plot too, with a tighter cluster near the left side of the graph and a noticeable spread farther to the right side. This suggests that the data is more heteroskedastic in nature. The p value below 0.05 in the var.test also lends itself towards the notion that the homoskedasticty assumption is not valid.

```
#qq test, sharpiro.test
qqnorm(reg2$residuals, cex=2)
qqline(reg2$residuals, col=2, lwd=3)
```



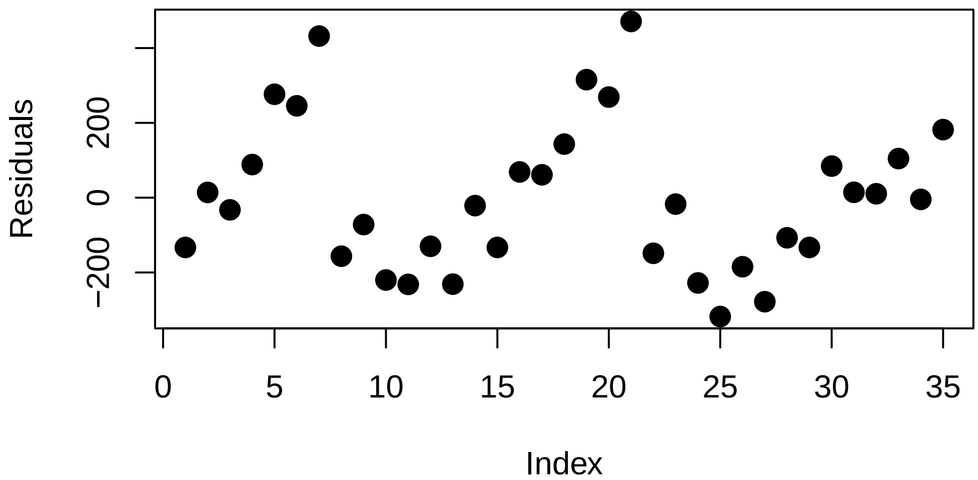

```
shapiro.test(reg2$residuals)
```

```
    Shapiro-Wilk normality test

data:  reg2$residuals
W = 0.9565, p-value = 0.1792
```

Normality

Compared to question 1, this QQ-plot has a much more noticeable drift from the line, with a large amount of drift above 1. Running the Shapiro-Wilk test for confirmation leads us to a large p-value of 0.1792, meaning we fail to reject the null hypothesis and assume normality is reasonably valid.

```r
plot(reg2$residuals, pch=16, cex=1.5, ylab="Residuals")
```



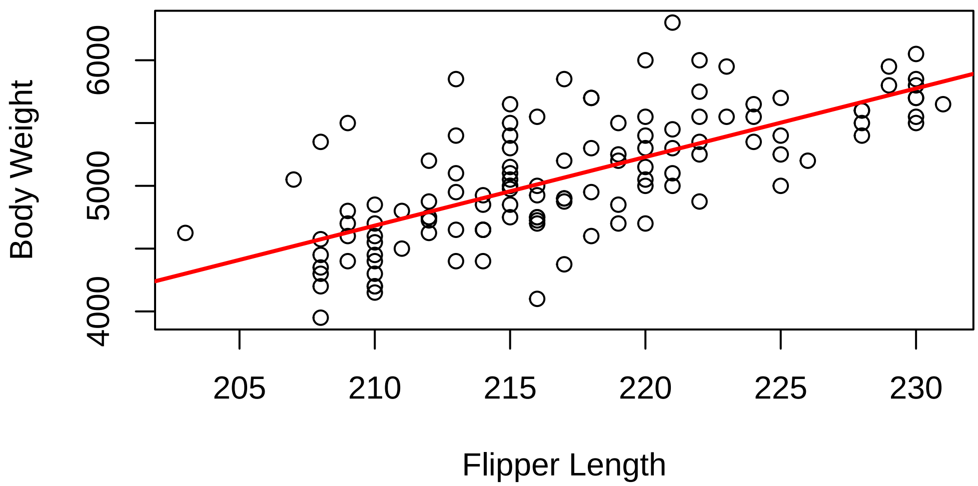

```r
dwtest(reg2)
```

```
	Durbin-Watson test

data:  reg2
DW = 0.95583, p-value = 0.0003126
alternative hypothesis: true autocorrelation is greater than 0
```

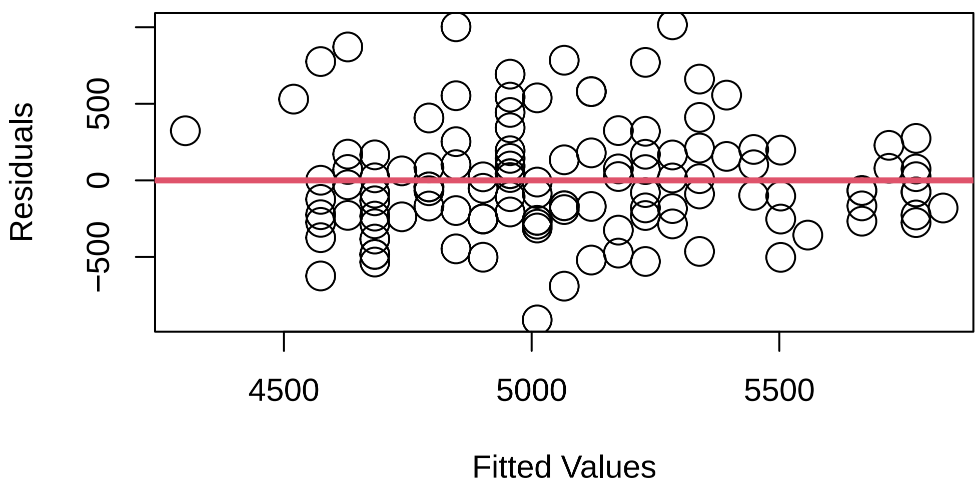
Independence

The residual plot shows a repeated, distinct wave-like pattern with three sections which suggests a violation of the independence assumption. If we examine it a little closer with a Durbin-Watson test we return a low p value of 0.0003, meaning we reject the null hypothesis of independence and our independence assumption is invalid.

Question 3



Body Weight
4000
5000
6000
205
210
215
220
225
230
Flipper Length



Residuals
−500
0
500
4500
5000
5500
Fitted Values

```
    F test to compare two variances

data:  reg3$residuals[group3] and reg3$residuals[!group3]
F = 0.8981, num df = 58, denom df = 63, p-value = 0.6804
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5412024 1.4984409
sample estimates:
ratio of variances
         0.8980979
```

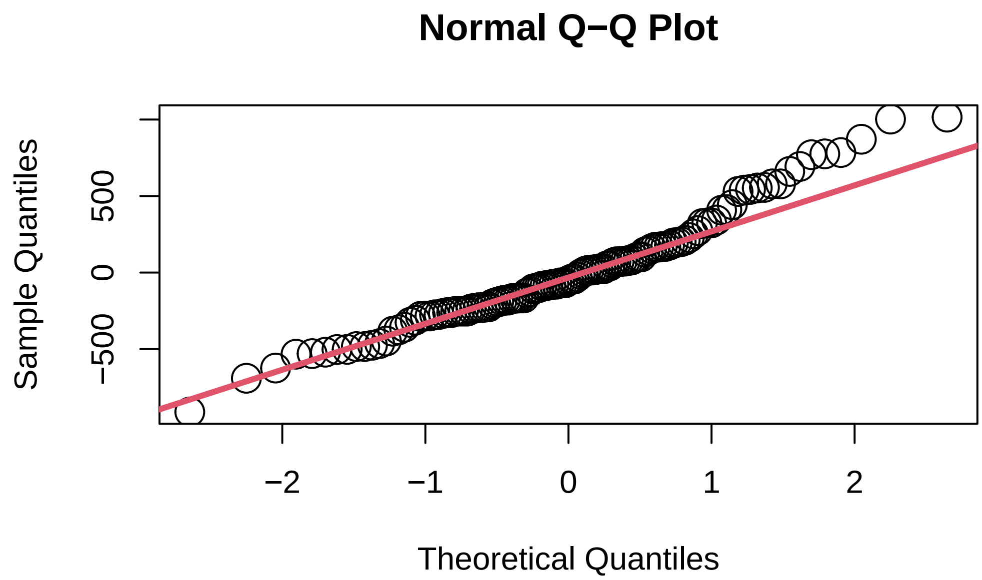
Linearity

Linearity appears valid in this data, as the fitted models shows a nice, even spread despite a decent amount of outliers that do not distort the overall trend. With our fitted vs residual plot, we get a random scatter of data with no visible curves or bends, which also supports the assumption of linearity in our data.

Homoskedasticity

The data in the fitted vs residual plot is tight, close to zero on the plot and does not have any visible fanning or spreading that suggests heteroskedasticity. The var test also yields favorable results with a ratio of variances at 0.8980979 and a p-value of 0.6804, meaning we fail to reject the null hypothesis and assume homoskedasticity.

```r
#qq test, sharpiro.test
qqnorm(reg3$residuals, cex=2)
qqline(reg3$residuals, col=2, lwd=3)
```



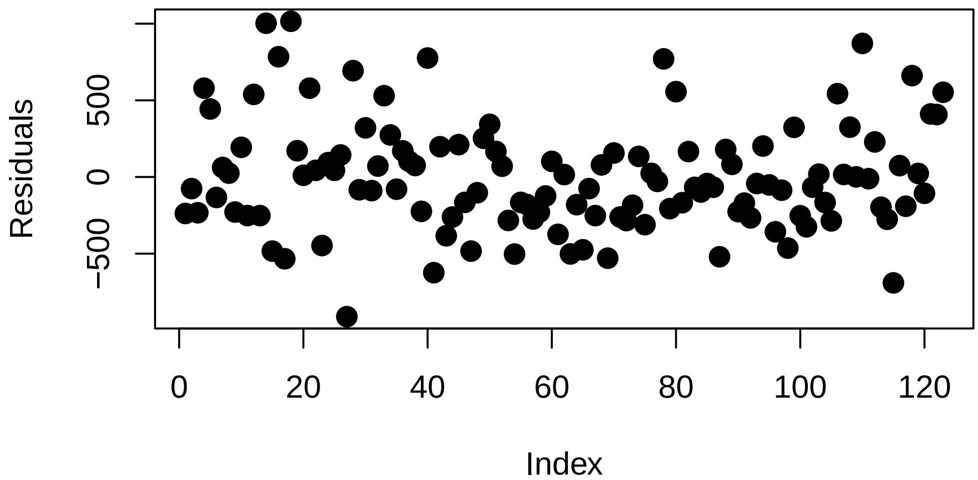

```r
shapiro.test(reg3$residuals)
```

```
    Shapiro-Wilk normality test

data:  reg3$residuals
W = 0.96845, p-value = 0.005573
```

Normality

There is a massive concentration of points centered around the line, so even the slight drift toward the edges initially did not deter me from assuming normality in this data. However, the Shapiro–Wilk test yields a p-value of 0.005573. Since this is below 0.05, we must reject the null hypothesis of normality. Despite the strong visual fit, the normality assumption is statistically invalid.

```r
plot(reg3$residuals, pch=16, cex=1.5, ylab="Residuals")
```




```r
dwtest(reg3)
```

```
	Durbin-Watson test

data:  reg3
DW = 2.395, p-value = 0.9875
alternative hypothesis: true autocorrelation is greater than 0
```

Independence

Unlike the previous dataset, there are no repeating patterns or waves in the residuals and the Durbin-Watson Test also yields a high p-value, all of which leads me to assume that independence in this dataset is a valid.