Ratio Variables

Andrew McAdam

2019-02-04

Contents

The Problems **1**

One Solution: ANCOVA **5**

A Better Solution: Power Law Models **10**

Error in X **13**

The Problems

Anytime you analyze a proportion, percentage or ratio variable as your response variable you are making some implicit assumptions that you might now be aware of! These assumptions can lead to some serious errors. So anytime you are tempted to analyze a ratio response variable (e.g., percent wet mass, captch per unit effort, etc) you should read through this document carefully!

I will use as an example a study by Packard and Boardman (1988). The paper is posted on CourseLink and they provided a dataset in this paper that is useful as an example.

```
painted<-read.table("data/painted.csv", sep=",", header=T)
painted$Group<-factor(painted$Group)
summary (painted)
```

```
##   CarcassMass     CarcassWater   Group
##  Min.   :2.492   Min.   :1.908   1:5
##  1st Qu.:2.850   1st Qu.:2.184   2:5
##  Median :3.372   Median :2.671
##  Mean   :3.381   Mean   :2.669
##  3rd Qu.:3.992   3rd Qu.:3.188
##  Max.   :4.264   Max.   :3.405
```

The goal of this study was to determine the effect of incubation humidity level (Group) on the water content of various organs (CarcassWater) in hatchling painted turtles, but the amount of water in the carcass is going to be influenced by the size of the hatchling turtle (CarcassMass). So you might be tempted to analyze these data by calculating the % water of the carcass (i.e., CarcassWater/CarcassMass*100) and to compare these percentages between the groups.

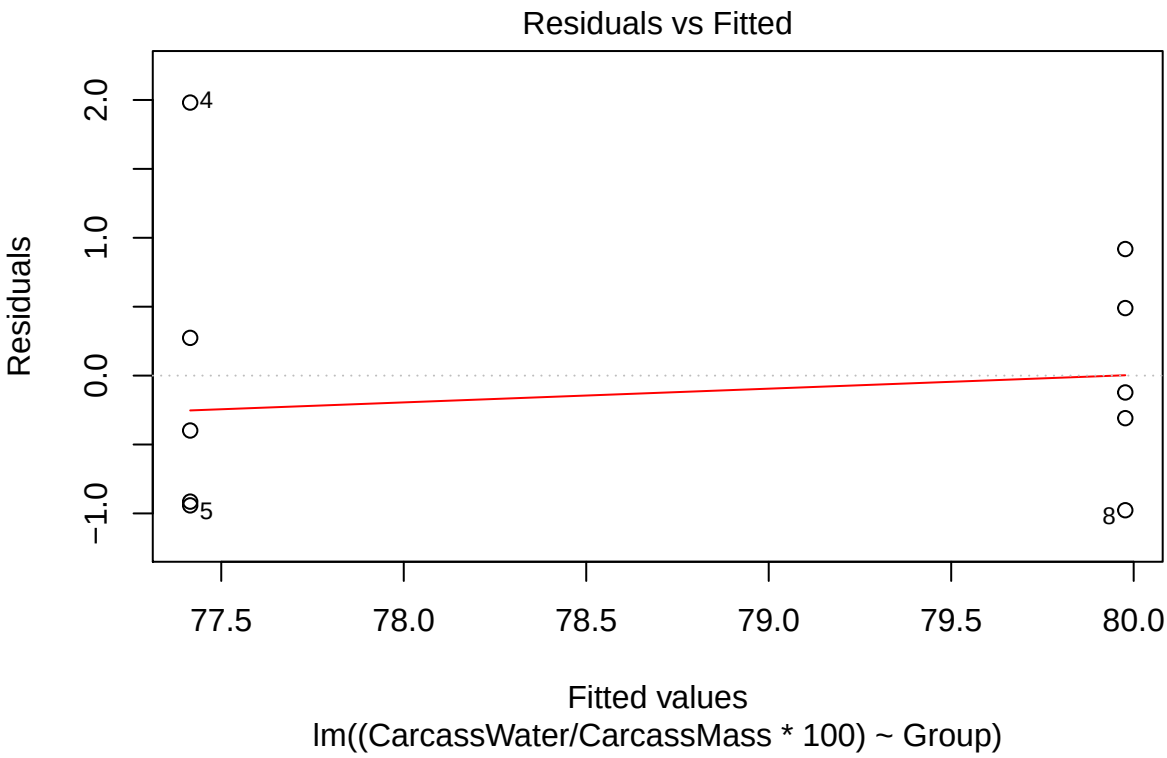
```
turtle.lm<-lm((CarcassWater/CarcassMass*100)~Group, data=painted)
summary (turtle.lm)
```

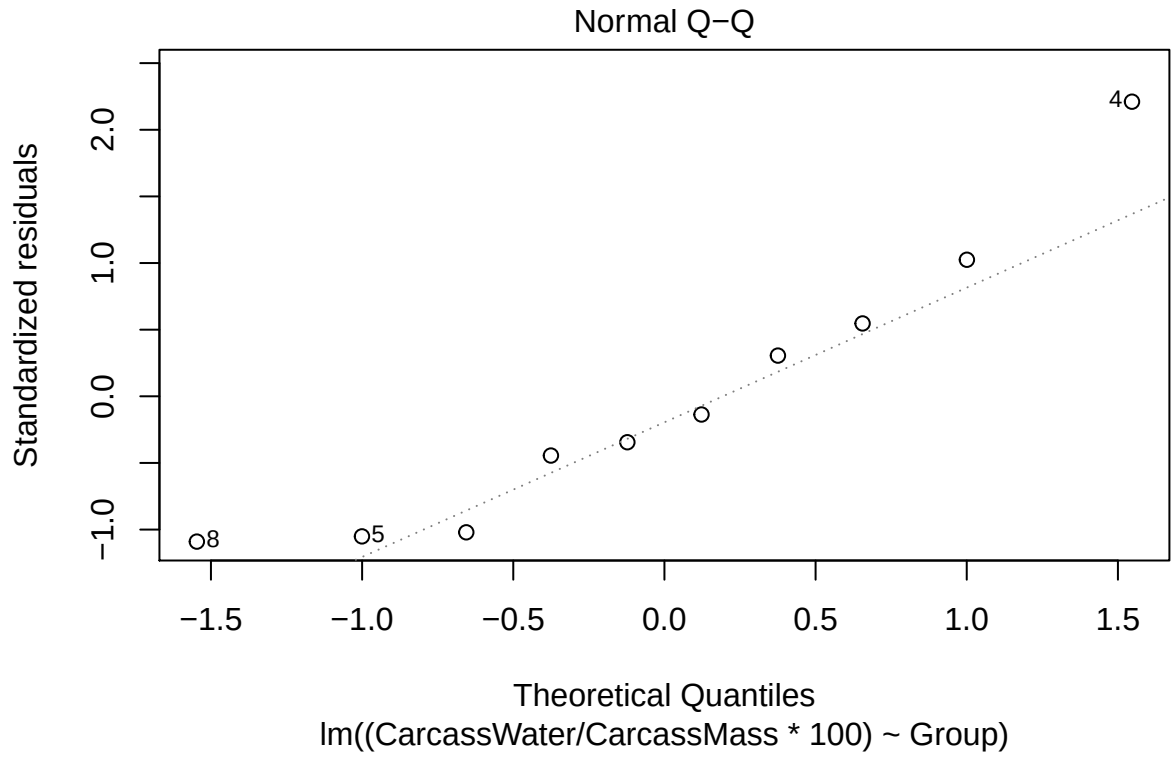
```
##
## Call:
## lm(formula = (CarcassWater/CarcassMass * 100) ~ Group, data = painted)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9770 -0.7856 -0.2156  0.4359  1.9812
```

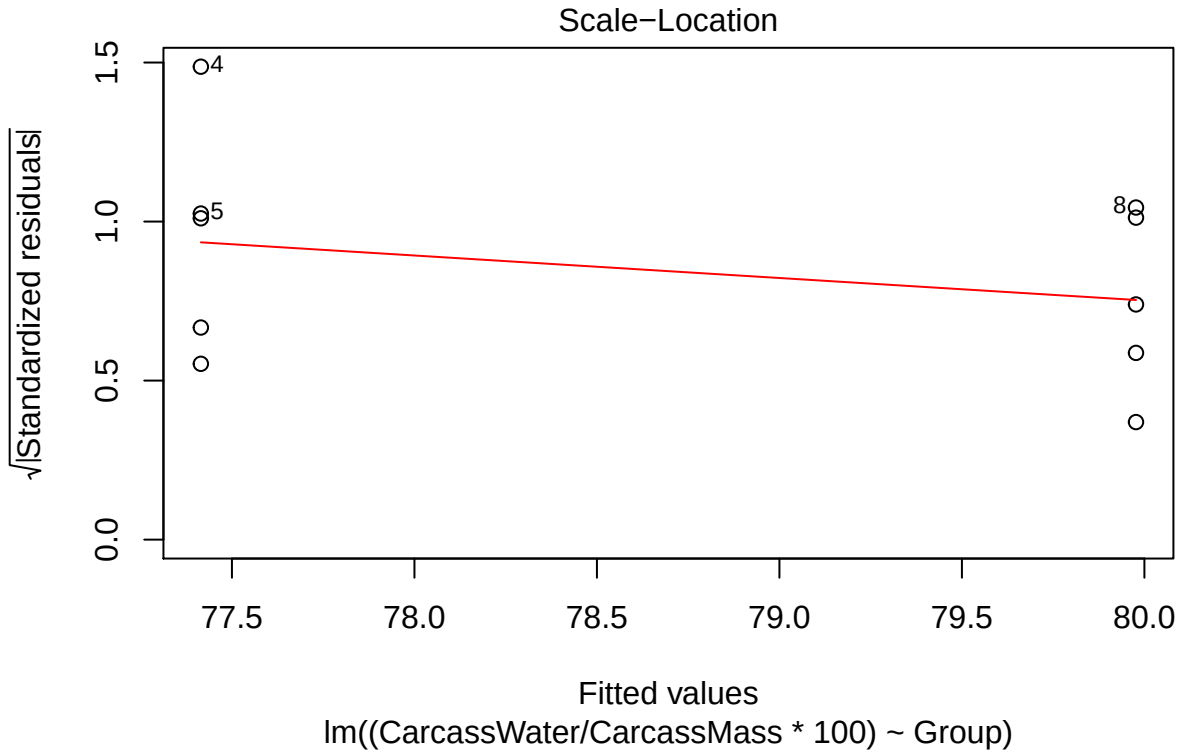
```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  77.4148     0.4481 172.774 1.41e-15 ***
## Group2        2.5622     0.6337   4.043  0.00372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.002 on 8 degrees of freedom
## Multiple R-squared:  0.6714, Adjusted R-squared:  0.6304
## F-statistic: 16.35 on 1 and 8 DF,  p-value: 0.003718
```

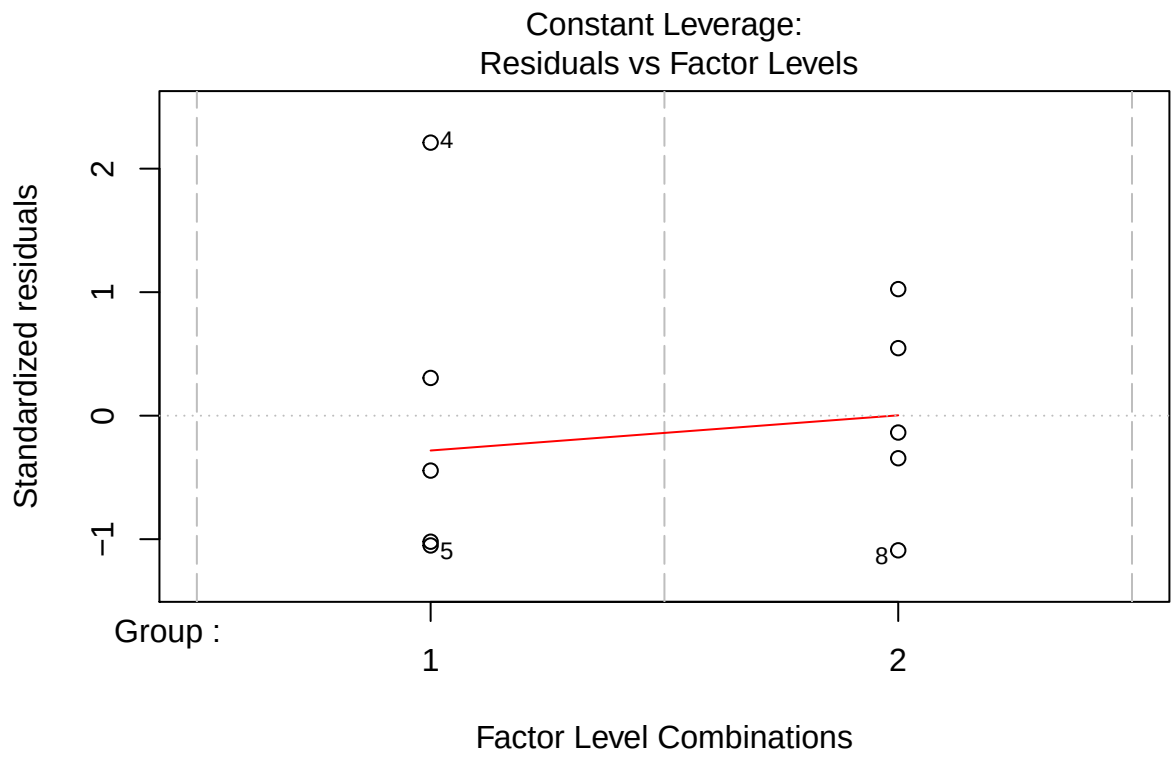
There are several problems with this approach. One is that a proportion is typically not normally distributed, especially when the proportion approaches 1 or 0. This can usually be accomodated by transformation or by using models that don't depend on normal errors. A more serious concern is that there is variation in the denominator (CarcassMass) and this variable may have responded to the treatment as well as the numerator, which is of primary interest. If the ratio depends on the value of the denominator, then the results could be confounded by the fact that the simple ratio does not adequately correct for the influence of carcass mass.

```r
plot(turtle.lm)
```

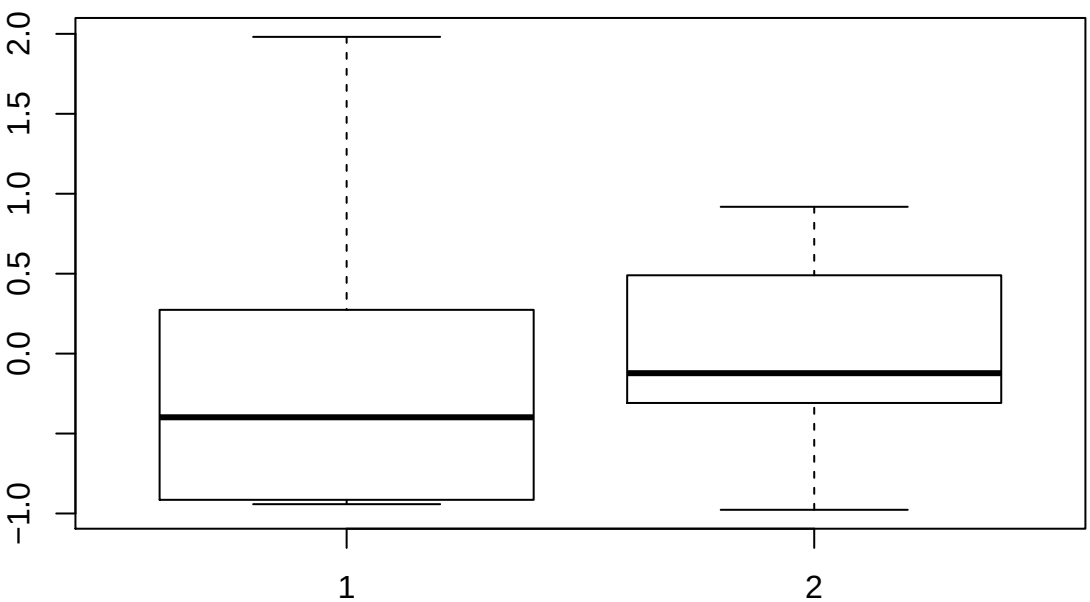



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm((CarcassWater/CarcassMass * 100) ~ Group)
4
5
8
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
−1.0
0.0
1.0
2.0

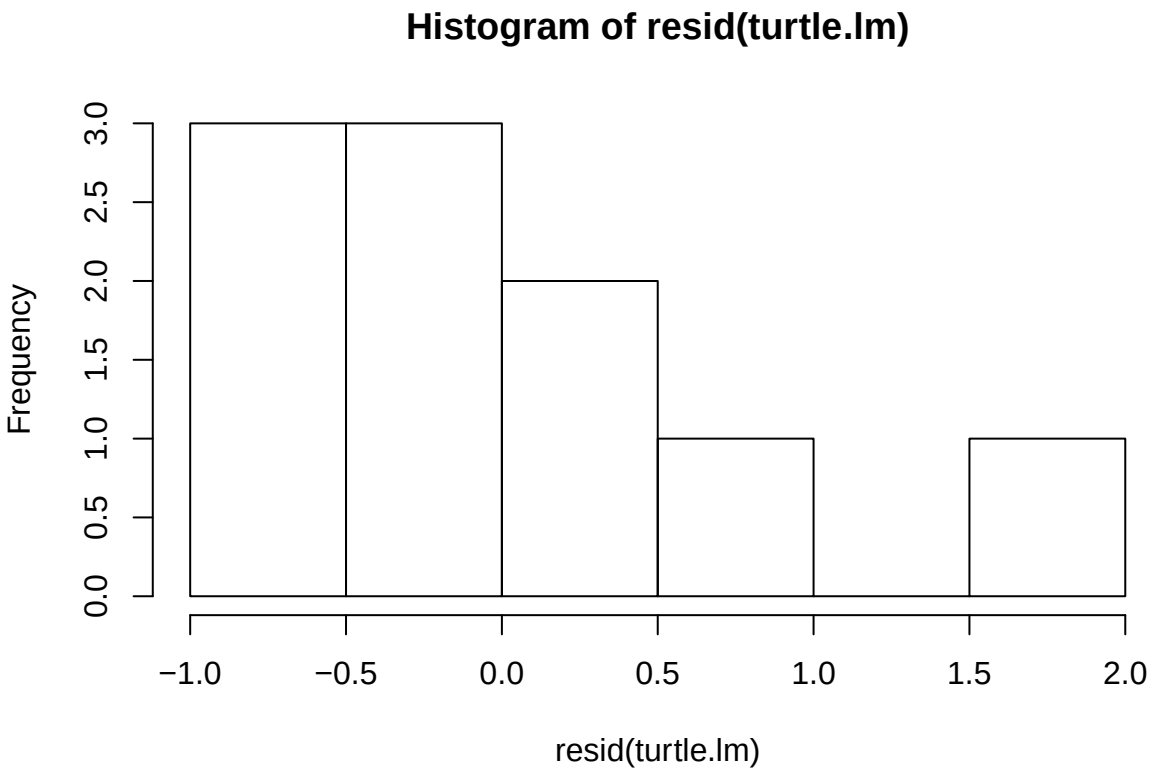

Scale−Location
√|Standardized residuals|
Fitted values
lm((CarcassWater/CarcassMass * 100) ~ Group)
4
5
8
77.5
78.0
78.5
79.0
79.5
80.0
0.0
0.5
1.0
1.5




```r
plot(painted$Group, resid(turtle.lm))
```




```r
hist(resid(turtle.lm))
```



More specifically analyzing a ratio as a response implicitly assumes that the ratio is isometric. That is to say that the proportional change in CarcassWater is associated with a proportional change in CarcassMass or alternatively that the % water is independent of CarcassMass.

One Solution: ANCOVA

A more appropriate approach to dealing with ratio responses is to instead use ANCOVA-like linear models

```
turtle.lm2<-lm(CarcassWater~CarcassMass+Group+CarcassMass:Group, data=painted)
summary (turtle.lm2)
```

```
##
## Call:
## lm(formula = CarcassWater ~ CarcassMass + Group + CarcassMass:Group,
##     data = painted)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027549 -0.020056 -0.004249  0.020293  0.029848
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.194618   0.091273  -2.132    0.077 .
## CarcassMass         0.843006   0.031479  26.780 1.79e-07 ***
## Group2              0.055335   0.160120   0.346    0.741
## CarcassMass:Group2 -0.007037   0.046099  -0.153    0.884
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.02774 on 6 degrees of freedom
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.9977
## F-statistic:  1292 on 3 and 6 DF,  p-value: 8.064e-09
```

In this case we test a series of hypotheses. First we test whether there is a difference between the groups in the relationship between CarcassWater and CarcassMass from the interaction term.

We can conclude from the above summary statement that this relationship does not differ between groups. We can therefore fit a model that assumes a common slope.

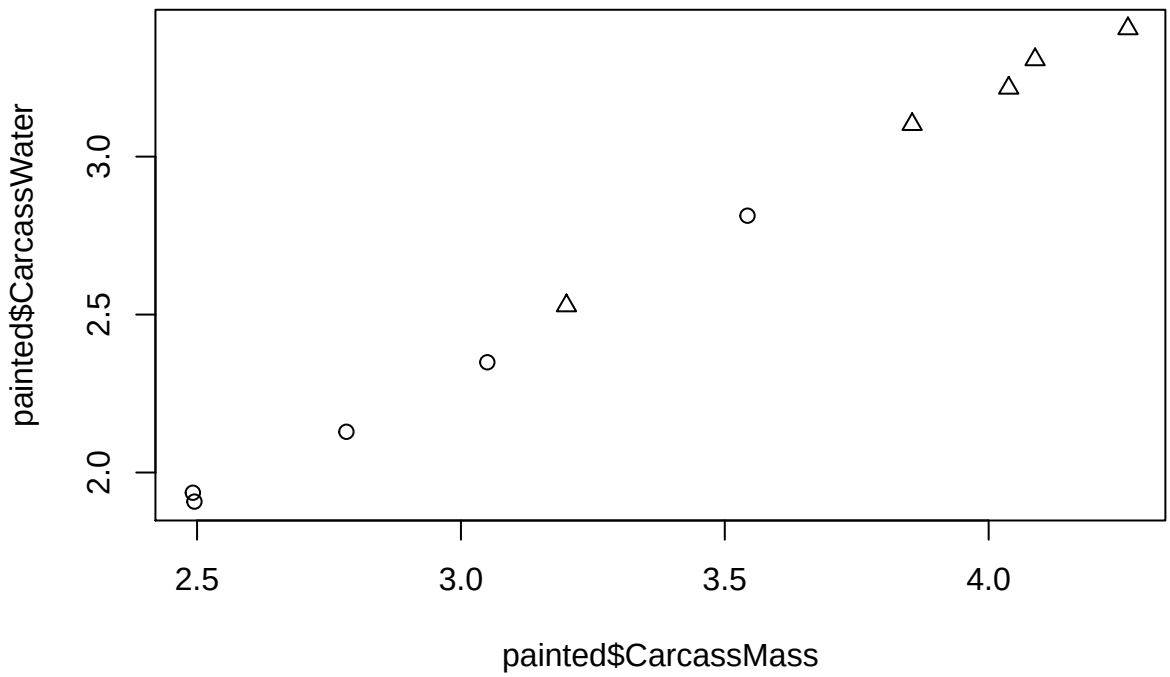
```r
turtle.lm3<-lm(CarcassWater~CarcassMass+Group, data=painted)
summary (turtle.lm3)
```

```
##
## Call:
## lm(formula = CarcassWater ~ CarcassMass + Group, data = painted)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026967 -0.021252 -0.003575  0.021974  0.028599
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.18519    0.06235  -2.970   0.0208 *
## CarcassMass  0.83972    0.02133  39.364 1.78e-09 ***
## Group2       0.03130    0.02711   1.155   0.2861
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02573 on 7 degrees of freedom
## Multiple R-squared:  0.9984, Adjusted R-squared:  0.998
## F-statistic:  2253 on 2 and 7 DF,  p-value: 1.47e-10
```

So there is indeed an effect of CarcassMass on CarcassWater as was initially suspected but once this is accounted for there in NO EFFECT of the treatment groups!!! This is a very different result from what we found previously. The reason for this is that CarcassWater and CarcassMass are not related isometrically to one another. That is THE % water IS NOT INDEPENDENT OF CARCASSMASS (the denominator) but instead is positively related to CarcassMass

It will be helpful to look at a few plots here

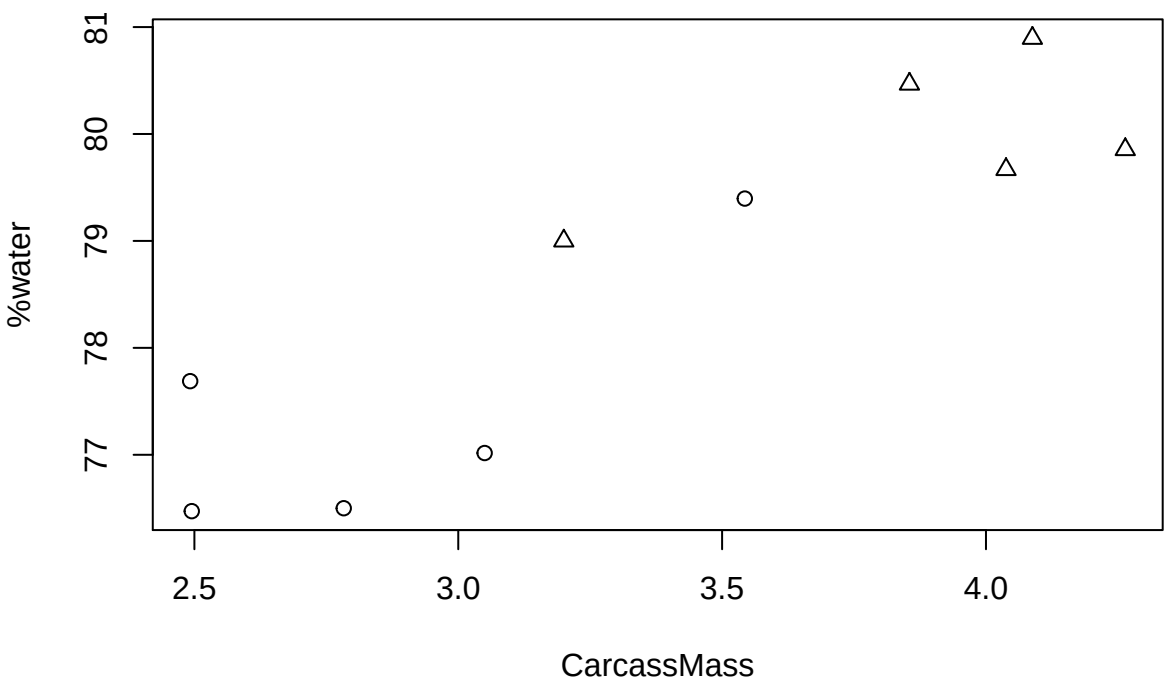
```r
plot(painted$CarcassMass, painted$CarcassWater, pch=as.numeric(painted$Group))
```



You can see from this plot that the points all fall along the same line and that there is clearly no effect of Group on the relationship. However, because the relationship is not isometric there was a difference found in the % water.

We can visualize the isometry relationship by...

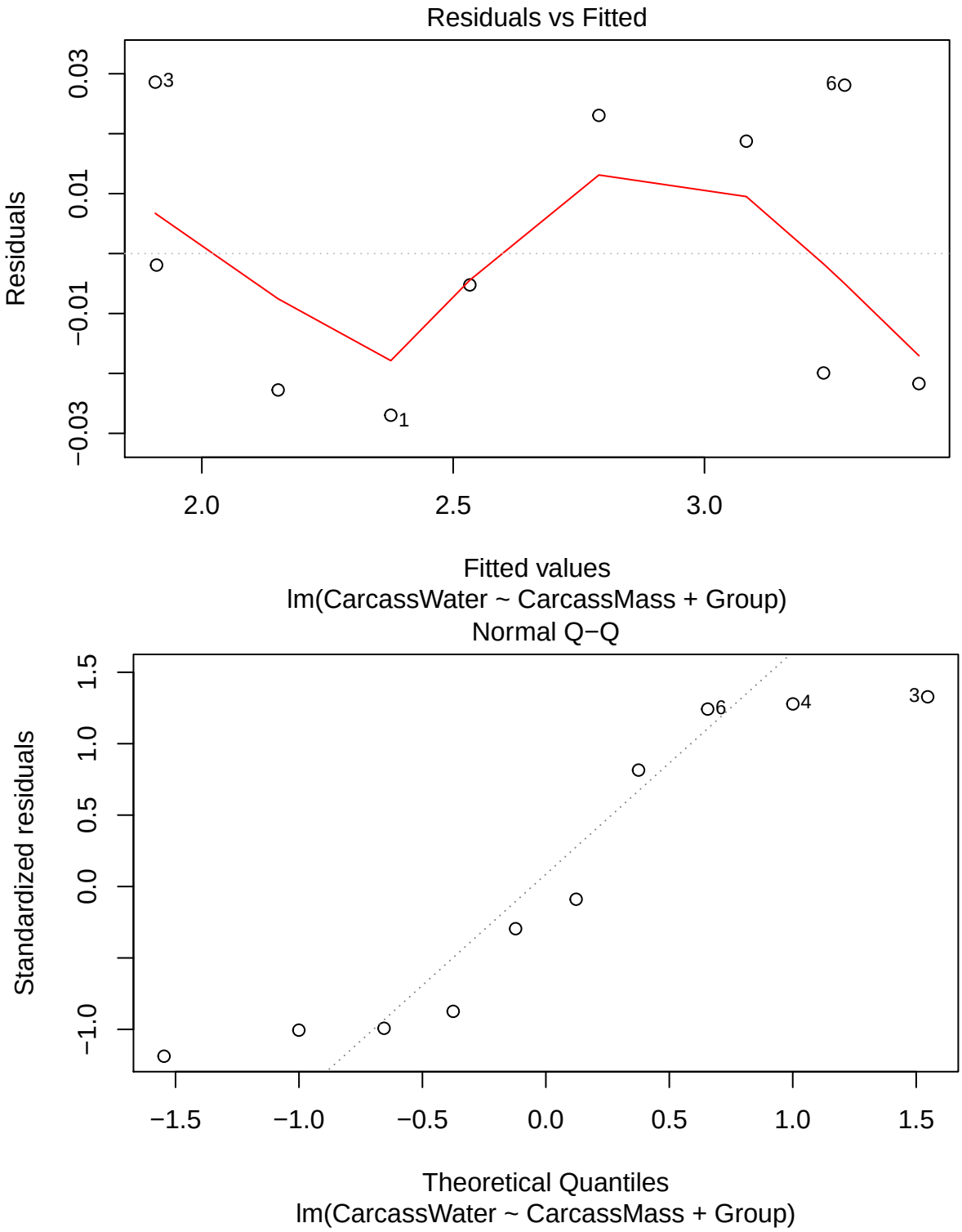
```
plot(CarcassWater/CarcassMass*100~CarcassMass, data=painted, pch=as.numeric(Group), ylab="%water")
```

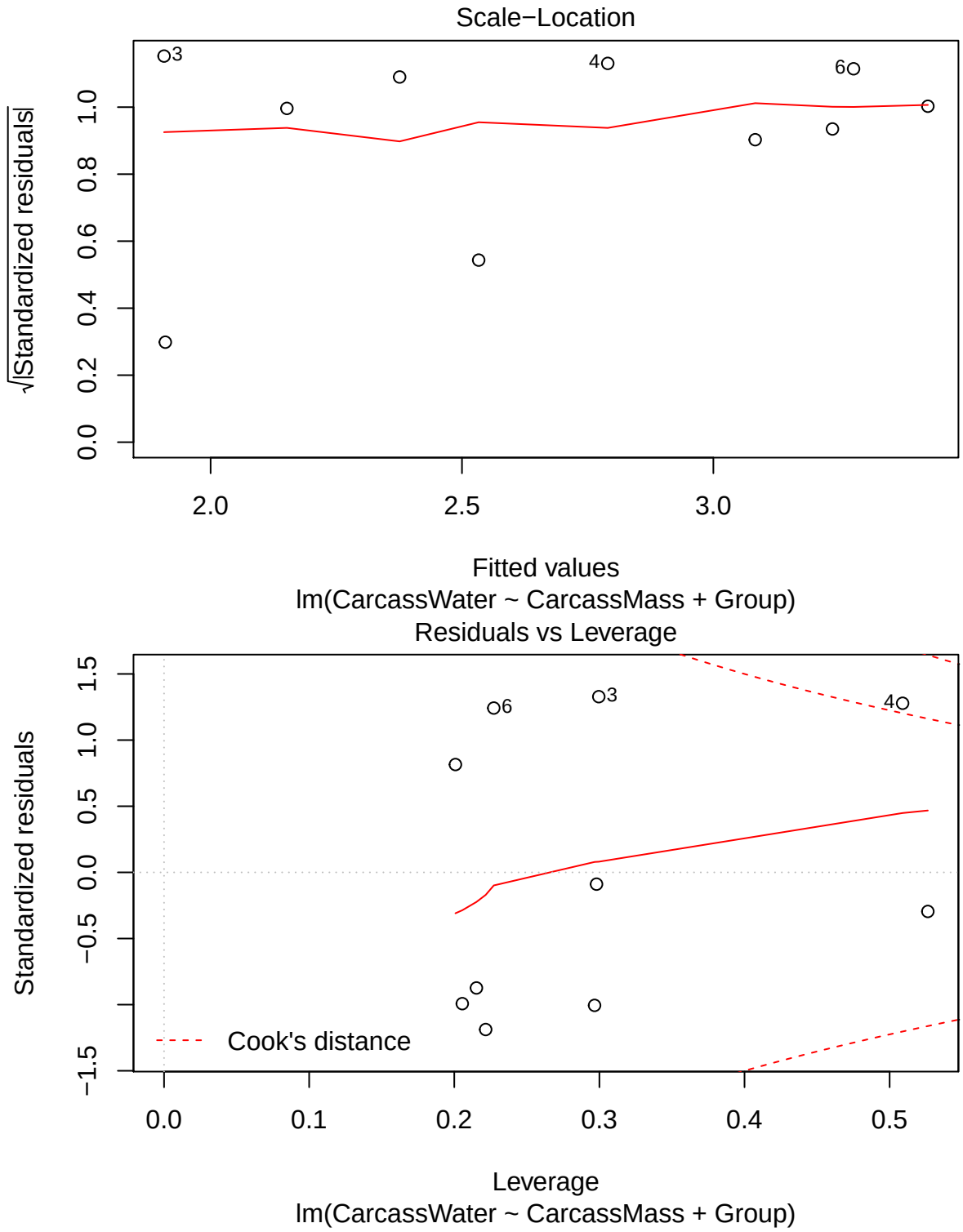


So the ratio is clearly not independent of the denominator!!!! The treatment effect on the ratio is confounded by the fact that there is still some of the treatment effect on the denominator that has not been accounted for by our simply % calculation. That is the ratio does not entirely account for the thing that it was supposed to account for!

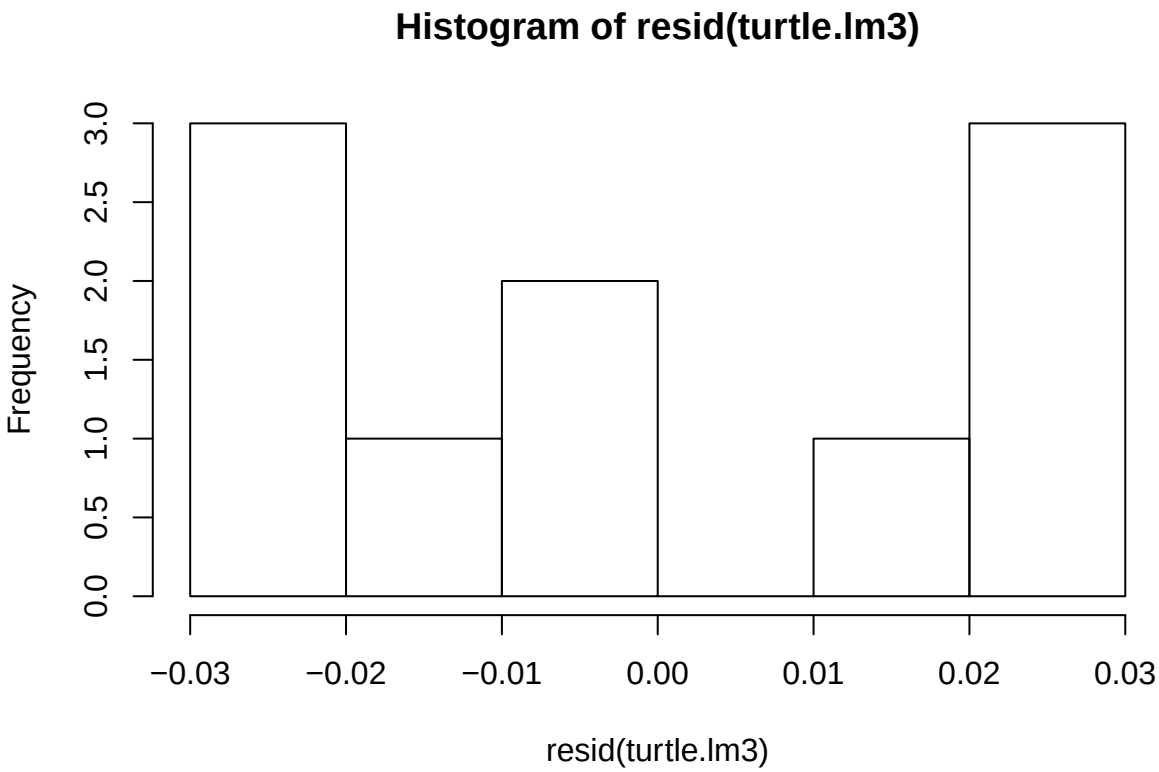
Packard and Boardman's ANCOVA approach correctly adjusts for the effect of CarcassMass on CarcassWater.

```
plot(turtle.lm3)
```

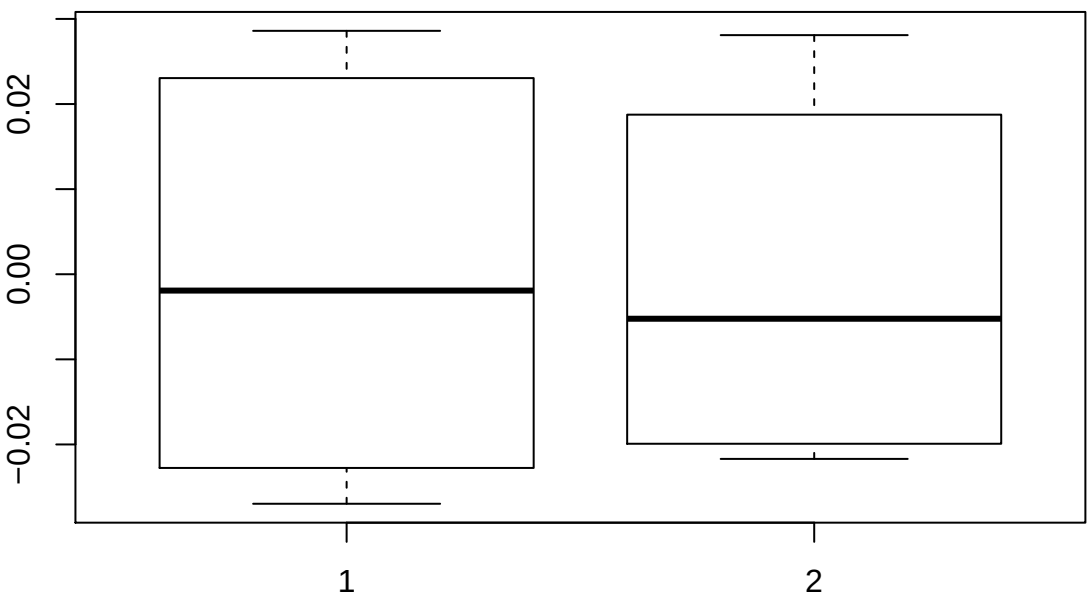





```
hist(resid(turtle.lm3))
```




```r
plot(painted$Group, resid(turtle.lm3))
```



The residuals of this model are still not great though.

A Better Solution: Power Law Models

We can further improve on Packard and Boardman's approach by recognizing that these two measures are related allometrically. As a result we can use a Power Law model to analyze these same data in a more transparent way. Recall that the Power Law equation is $Y = Ax^b$, which can be linearized by log transforming

both sides of the equation:

$$log(y) = a + blog(x),$$

where a = log(A).

```
turtle.lm4<-lm(log(CarcassWater)~log(CarcassMass)+Group+log(CarcassMass):Group, data=painted)
summary (turtle.lm4)
```

```
##
## Call:
## lm(formula = log(CarcassWater) ~ log(CarcassMass) + Group + log(CarcassMass):Group,
##     data = painted)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.010208 -0.006424 -0.002321  0.008034  0.013563
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.33418    0.03769  -8.867 0.000114 ***
## log(CarcassMass)         1.07464    0.03574  30.072 8.97e-08 ***
## Group2                   0.04016    0.07444   0.539 0.608964
## log(CarcassMass):Group2 -0.02251    0.05929  -0.380 0.717303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01062 on 6 degrees of freedom
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.9977
## F-statistic:  1320 on 3 and 6 DF,  p-value: 7.575e-09
```

Again the interaction is not significant so the relationship is the same in both groups. We can remove it and analyze the simpler model in which there is a common slope in the two groups.

```
turtle.lm5<-lm(log(CarcassWater)~log(CarcassMass)+Group, data=painted)
summary (turtle.lm5)
```

```
##
## Call:
## lm(formula = log(CarcassWater) ~ log(CarcassMass) + Group, data = painted)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.010281 -0.007618 -0.001498  0.007459  0.012475
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.32562    0.02830 -11.506 8.43e-06 ***
## log(CarcassMass)  1.06646    0.02671  39.921 1.61e-09 ***
## Group2            0.01221    0.01034   1.181    0.276
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009946 on 7 degrees of freedom
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.998
## F-statistic:  2255 on 2 and 7 DF,  p-value: 1.464e-10
```

Our interpretation of the results are the same as above, but this approach has an added benefit. In this

approach the slope above (1.066) has some added information and can be used to test against some interesting theoretical expectations. Recall that the Power Law model is:

$$CarcassWater = A * CarcassMass^b$$

These results indicate that $log(CarcassWater) = -0.33 + 1.066log(CarcassMass) + 0.012(Group2?)$, which can be back-transformed to the nonlinear form as:

$$CarcassWater = exp(-0.33 + 0.012(Group2?)) * CarcassMass^{1.066}$$

This is useful because we can examine the relationship between the proportion of water and the CarcassMass

$$CarcassWater/CarcassMass = exp(-0.33 + 0.012(Group2?)) * CarcassMass^{0.066}$$

Note that the slope (b) above is slightly greater than zero (0.066) so the proportion of water is positively associated with CarcassMass. If this adjusted slope was equal to zero then the proportion of water would be indepedent of CarcassMass (i.e. a constant) and the relationship would be isometric. This would mean that it would be ok to simply compare the proportion between groups (but note the problems with the residuls).

Note that the effects of Group on the proportion is also clear in this notation! Some people worry about losing this information because the proportion has some common interpretation (e.g., catch per unit effort).

The benefit of this approach is that we can directly test the null hypothesis that the proportion does not change with CarcassMass as would be indicated by a slope of one in the original model (zero in the model with the proportion).

We can put 95% confidence intervals on the slope using:

```r
summary (turtle.lm5)
```

```
##
## Call:
## lm(formula = log(CarcassWater) ~ log(CarcassMass) + Group, data = painted)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.010281 -0.007618 -0.001498  0.007459  0.012475
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.32562    0.02830 -11.506 8.43e-06 ***
## log(CarcassMass)  1.06646    0.02671  39.921 1.61e-09 ***
## Group2            0.01221    0.01034   1.181    0.276
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009946 on 7 degrees of freedom
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.998
## F-statistic:  2255 on 2 and 7 DF,  p-value: 1.464e-10
```

95% CL = slope+/- qt(0.975, df=7)*se

```r
1.06646 +(qt(0.975, df=7)*0.02671)
```

```
## [1] 1.129619
```

```r
1.06646 -(qt(0.975, df=7)*0.02671)
```

```
## [1] 1.003301
```

So the 95% CL are (1.003, 1.1296). The parameter therefore significantly exceeds 1 so the relationship is not isometric.

Note that diagnostic plots still show some problems with this model, but we will not deal with that here.

Error in X

Recall that one of our assumptions of a general linear model is that all of the unexplained variation (error) is in the response variable and the predictor variables are measured without error (see the glm document). In this case we clearly have error in a predictor variable AND we are interested in the value of the slope and not just its significance (i.e. differnce from zero). So in this case it would likely be important to use a TYPE-II regression to test whether the slope is different from 1.