Mahidol University

Abalone

Programming in statistics

Suttada | Rasita | Vipoo
Ponlapee | Atcharawan

R
Programming

Overview

What is Abalone?

what is Abalone

หอยเป๋าฮื้อ

อาจเรียกว่า หอยร้อยรู หรือ หอยโข่งทะเล เป็นหอยทะเลฝาเดียว
อยู่ในกลุ่ม มอลลัสค์ (mollusk) ซึ่งอยู่ในไฟลัม mollusca
หอยเป๋าฮื้อมีรสชาติอร่อย ไม่คาวจัด มีโปรตีนสูง ไขมันต่ำ
ให้แคลอรีต่ำ นิยมรับประทานกันอย่างกว้างขวาง และมีราคาสูง

Dataset details

01

4177 ข้อมูล 9 ตัวแปร

น้ำหนักทั้งหมดของหอย

น้ำหนักเฉพาะส่วนเนื้อหอย

น้ำหนักส่วนอวัยวะภายใน

น้ำหนักเฉพาะส่วนเปลือกหอย

sex	Length (mm)	Diameter (mm)	Height (mm)	Whole. weight (g)	Shucked. weight (g)	Viscera. weight (g)	Shell. weight (g)	rings
F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7


CALMEX
Shucked
weight

Abalone Viscera

มีส่วนช่วยป้องกันการติดไวรัส โควิด-19

marine drugs MDPI

Article

Inhibition of SARS-CoV-2 Virus Entry by the Crude Polysaccharides of Seaweeds and Abalone Viscera In Vitro

Sung-Kun Yim [1,*], Kian Kim [1], In-Hee Kim [2], Sang-Ho Chun [1], Tae-Hwan Oh [1], Jin-Ung Kim [3], Jung-Won Kim [2], Woo-Huk Jung [1], Ho-Sang Moon [2], Bo-Sung Ku [2] and Kyoo-Jin Jung [1]

1 Marine Biotechnology Research Center, Jeonnam Bioindustry Foundation, 21-7, Nonggongdanji 4Gil, Wando-eup, Wando-gun, Jeollanam-do 59108, Korea; kimka3152@jbf.kr (K.K.); csh@jbf.kr (S.-H.C.); sosoth@jbf.kr (T.-H.O.); woohyuki@jbf.kr (W.-H.J.); kjjung@jbf.kr (K.-J.J.)
2 Medical & Bio Decision (MBD) Co. Ltd., #B-8F, 145 Gwanggyo-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16229, Korea; ihkim@mbdbiotech.com (I.-H.K.); jungwonk@mbdbiotech.com (J.-W.K.); novagene@mbdbiotech.com (H.-S.M.); goos4684@mbdbiotech.com (B.-S.K.)
3 Marine Bio Team, Wando County, 51, Chonghaejinnam-ro, Wando-eup, Wando-gun, Jeollanam-do 59124, Korea; gadget21c@korea.kr
* Correspondence: skyim01@jbf.kr; Tel.: +82-10-4016-6745; Fax: +82-61-555-1797

Viscera

ANOVA 02

1 **Shucked Weight**

ตัวแปรต้น : Sex (F,M,I)
ตัวแปรตาม : Shucked Weight

2 **Viscera Weight**

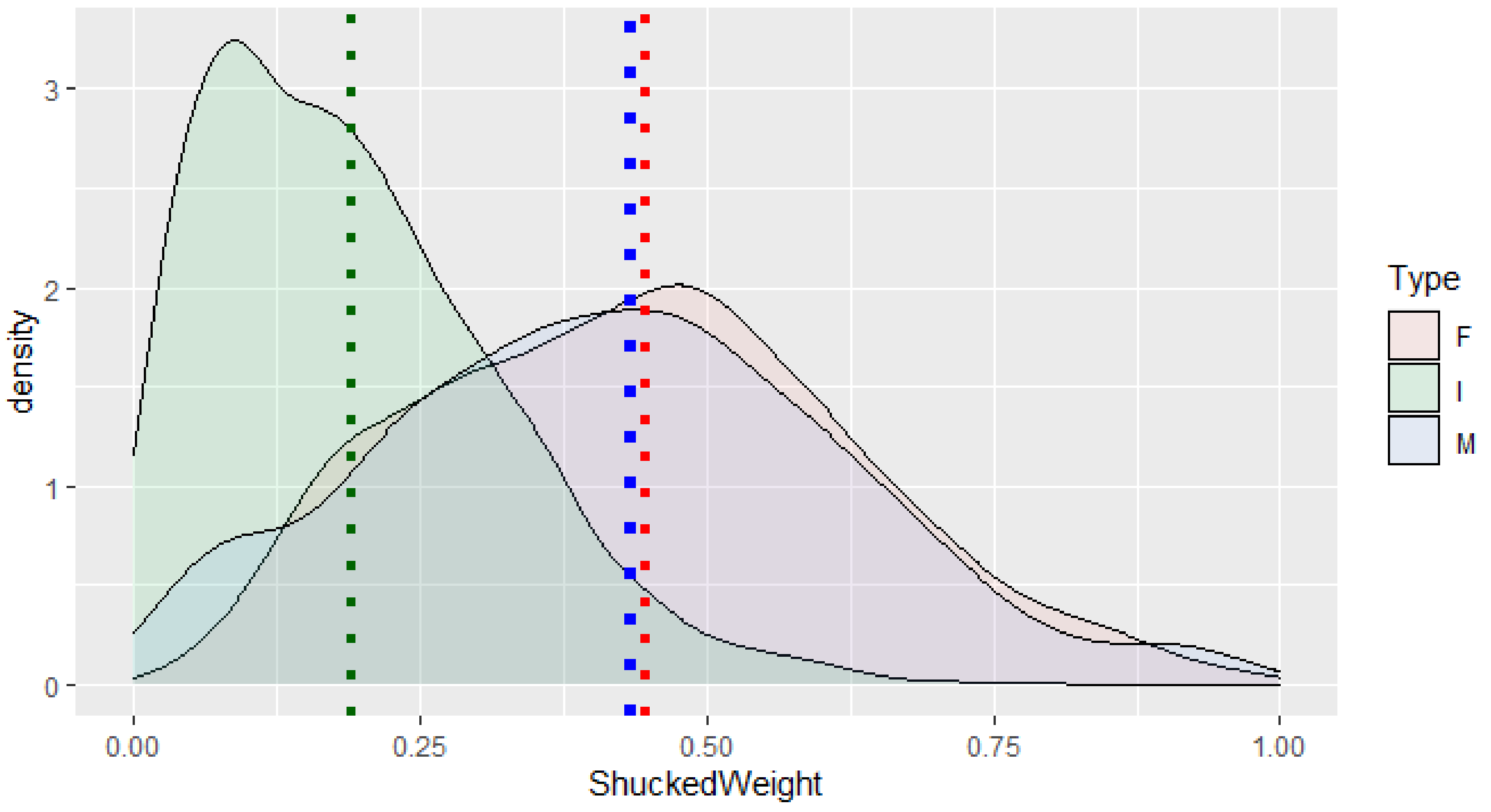
ตัวแปรต้น : Sex (F,M,I)
ตัวแปรตาม : Viscera Weight

- การหาความแตกต่างของค่าเฉลี่ยน้ำหนักส่วนเนื้อหอยในแต่ละเพศ
- การหาความแตกต่างของค่าเฉลี่ยน้ำหนักส่วนอวัยวะภายในหอยในแต่ละเพศ

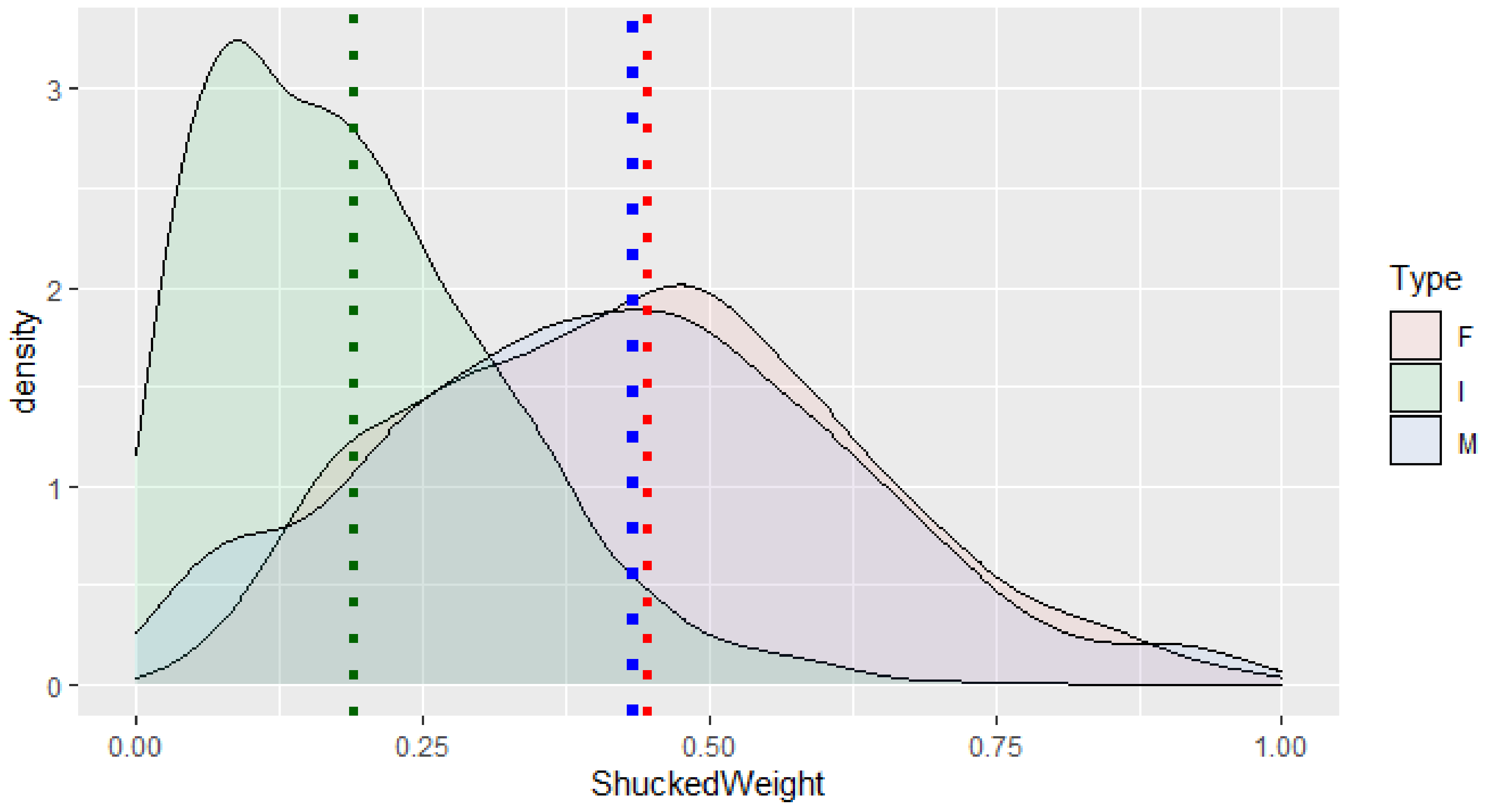
Shucked Weight

ANOVA

01 w/ outlier



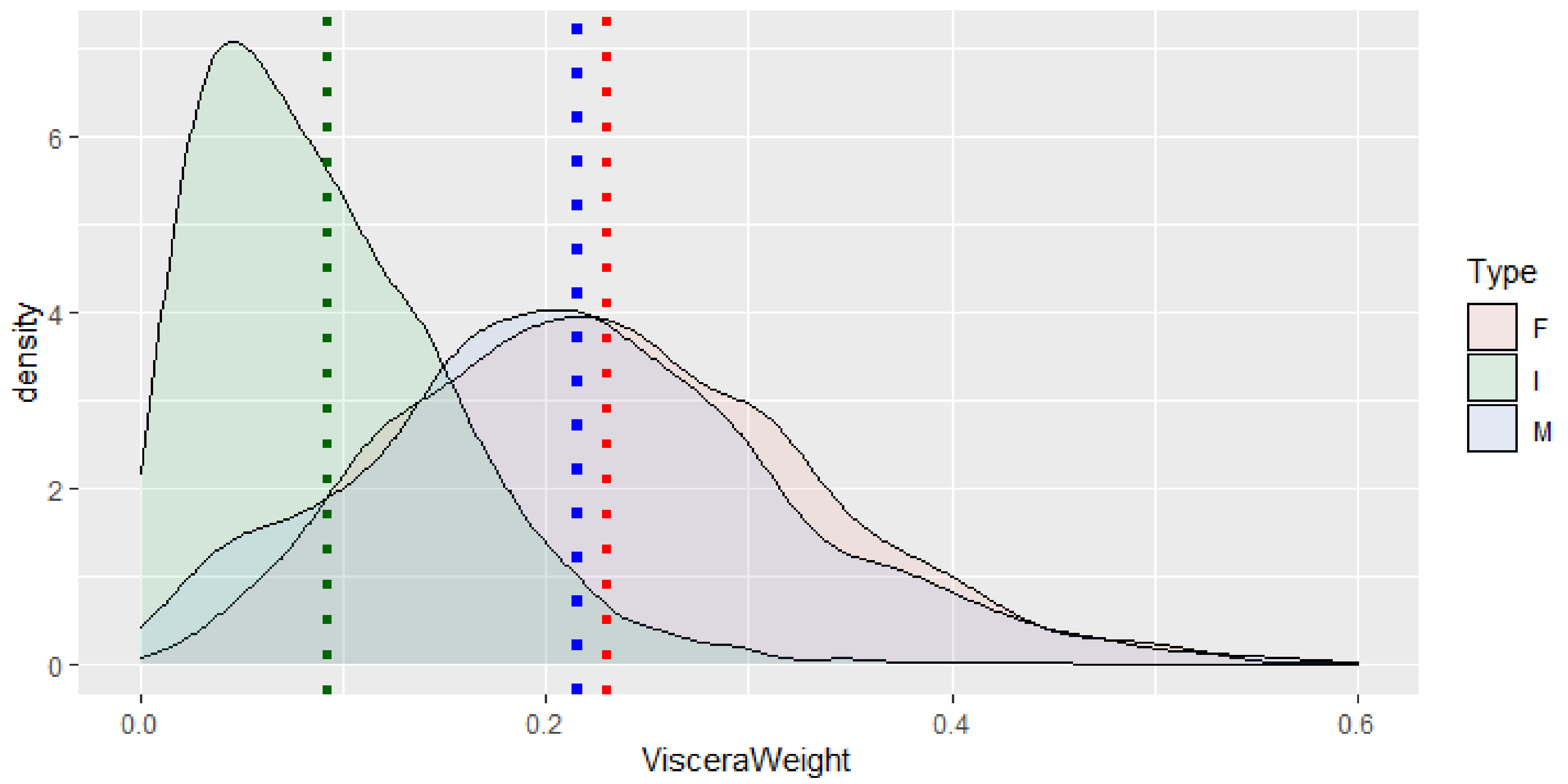
02 w/o outlier



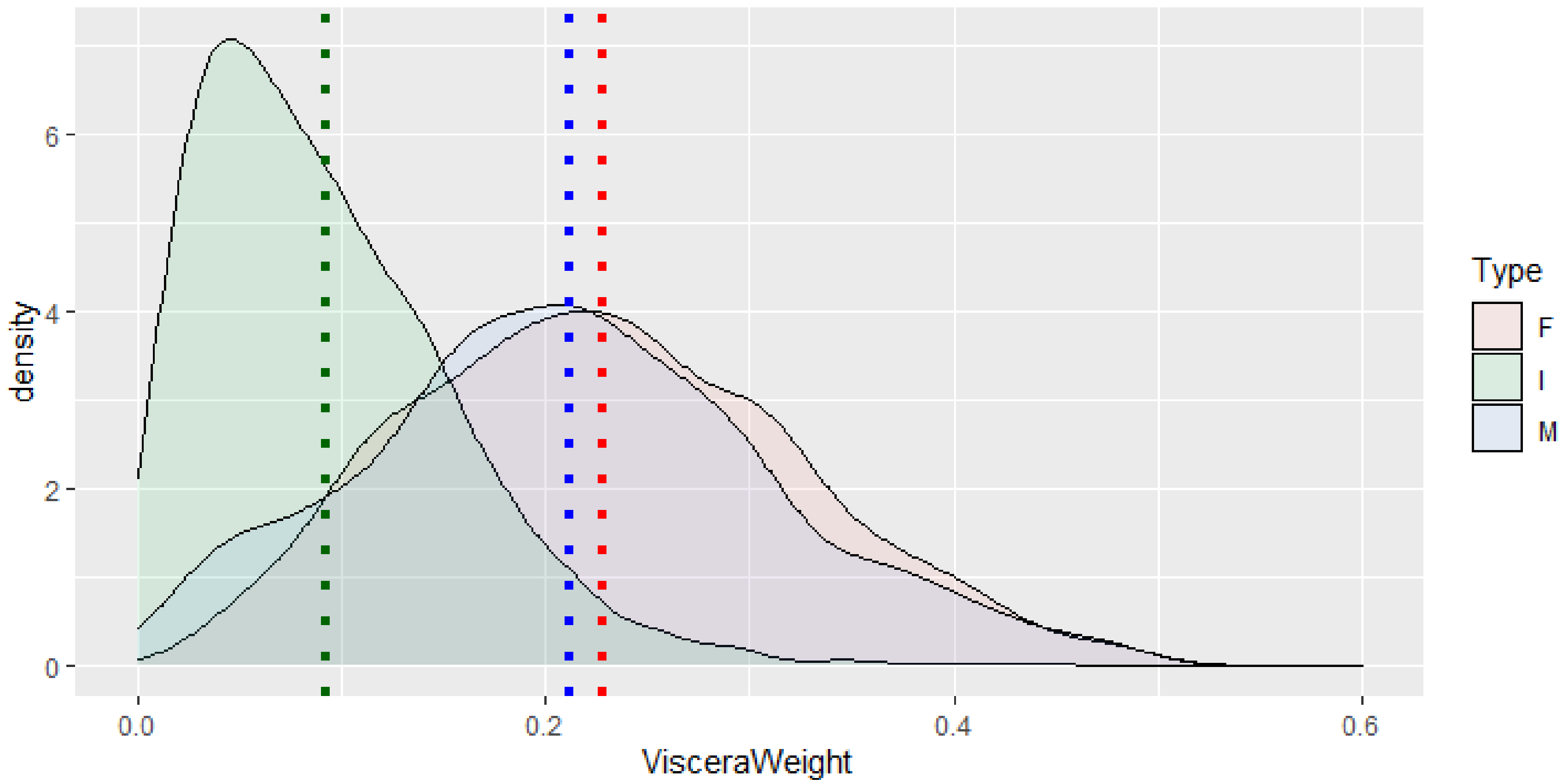
Viscera Weight

ANOVA

01 w/ outlier



02 w/o outlier



ANOVA Assumptions

Shucked Weight

01 Equal Variances

```
> abalone_shuckedweight.sd <- tapply(abalone$ShuckedWeight, INDEX = abalone$Type, FUN = sd)
> abalone_shuckedweight.sd
        F         I         M 
0.1986632 0.1284053 0.2230000 
```

```
> abalonesh.anova <- aov(ShuckedWeight~Type,data = abalone)
> summary(abalonesh.anova)
> leveneTest(abalonesh.anova)

Levene's Test for Homogeneity of Variance (center = median)
        Df F value    Pr(>F)    
group    2     142 < 2.2e-16 ***
      4174                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> max(abalone_shuckedweight.sd)/min(abalone_shuckedweight.sd)
[1] 1.736688
```

02 Normality

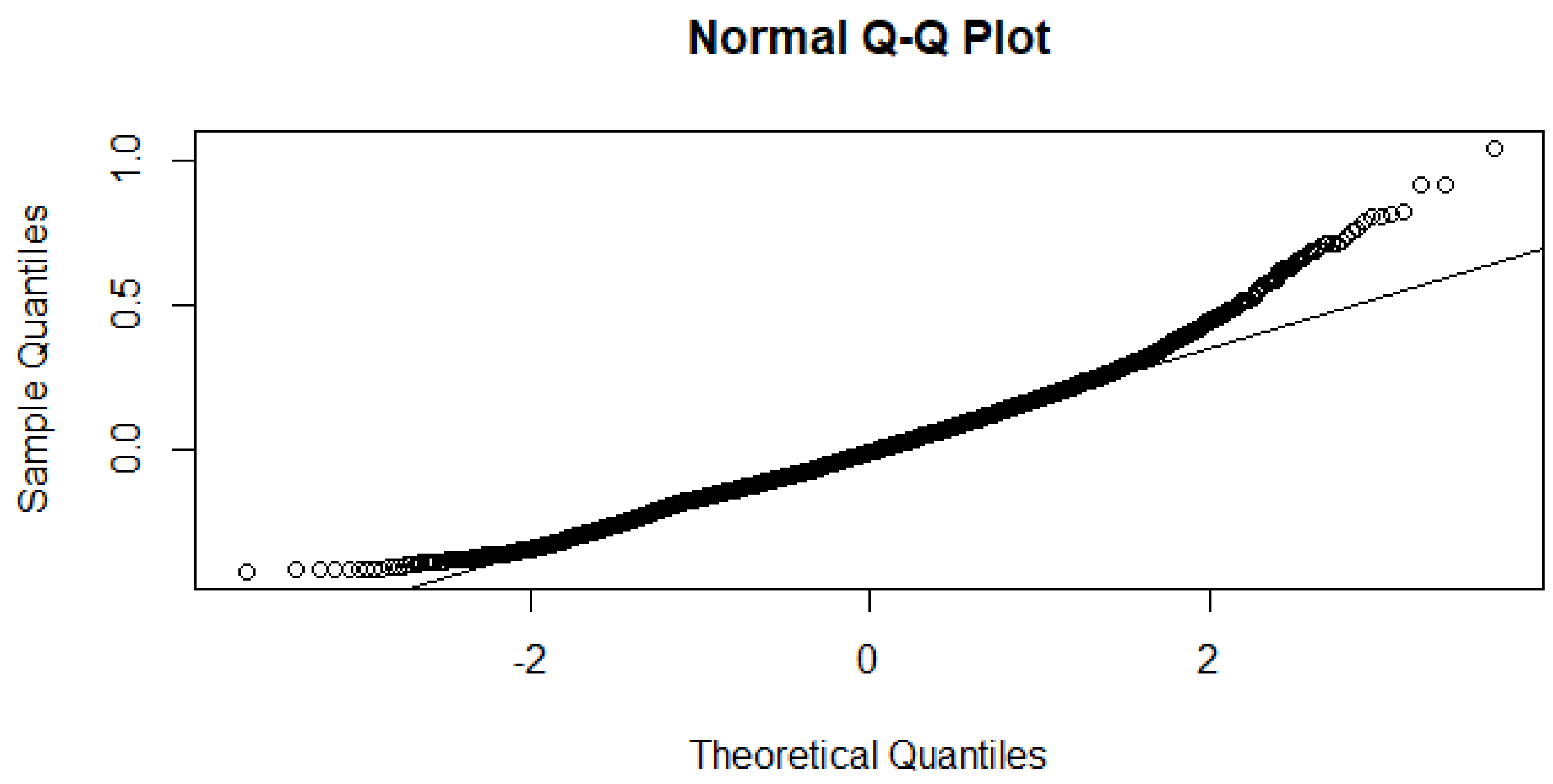
```
> model  <- lm(ShuckedWeight ~ Type, data = abalone)
> shapiro.test(residuals(model))

	Shapiro-Wilk normality test

data:  residuals(model)
W = 0.9765, p-value < 2.2e-16
```

03 Independence

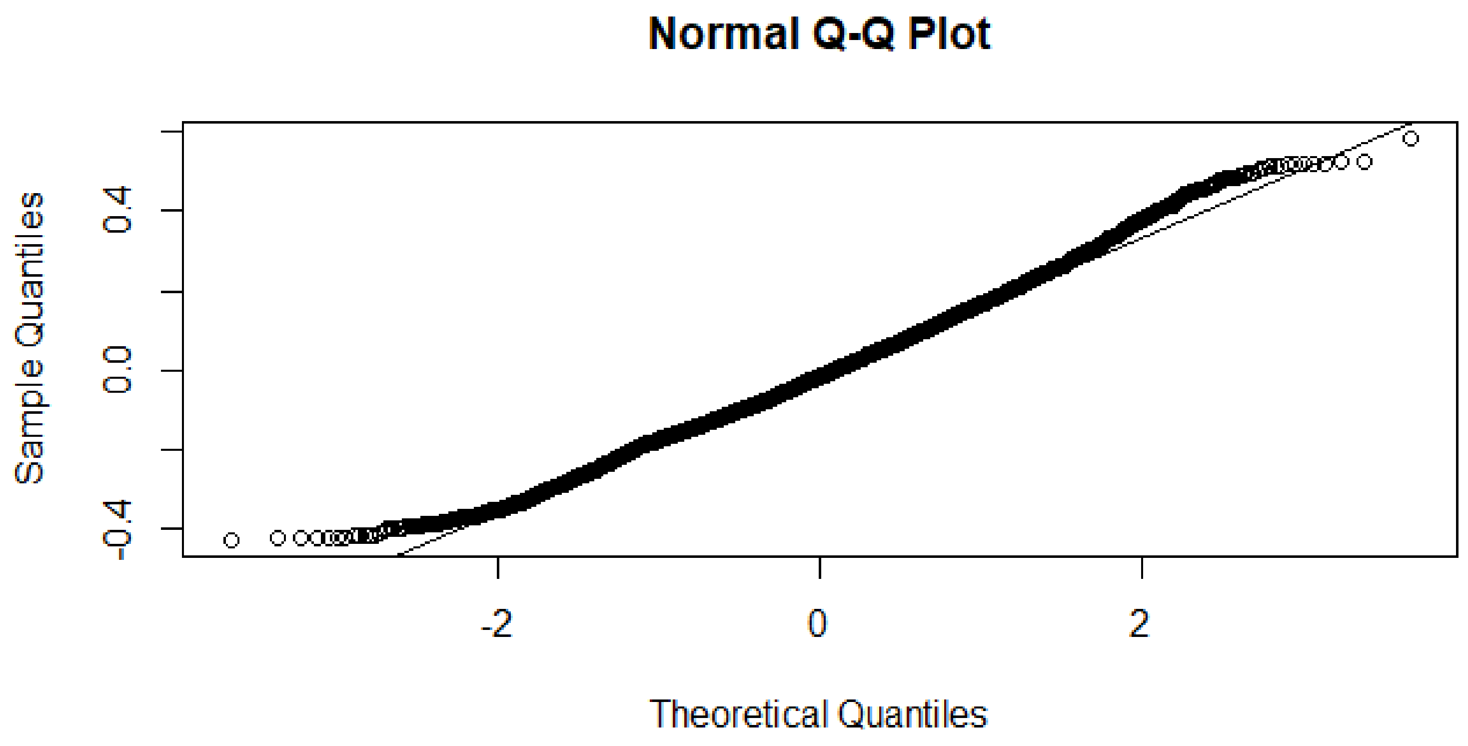
Shucked Weight

ANOVA

01 w/ outlier



02 w/o outlier



ANOVA Assumptions

Viscera Weight

01 Equal Variances

```
> abalone.vi.sd <- tapply(abalone$VisceraWeight, INDEX = abalone$Type, FUN = sd)
> abalone.vi.sd
         F          I          M
0.09761738 0.06253609 0.10491899
```

```
> abalonevw.anova <- aov(VisceraWeight~Type,data = abalone)
> summary(abalonevw.anova)
> leveneTest(abalonevw.anova)

Levene's Test for Homogeneity of Variance (center = median)
        Df F value    Pr(>F)
group    2  142.49 < 2.2e-16 ***
      4174
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> max(abalone.vi.sd)/min(abalone.vi.sd)
[1] 1.677735
```

02 Normality

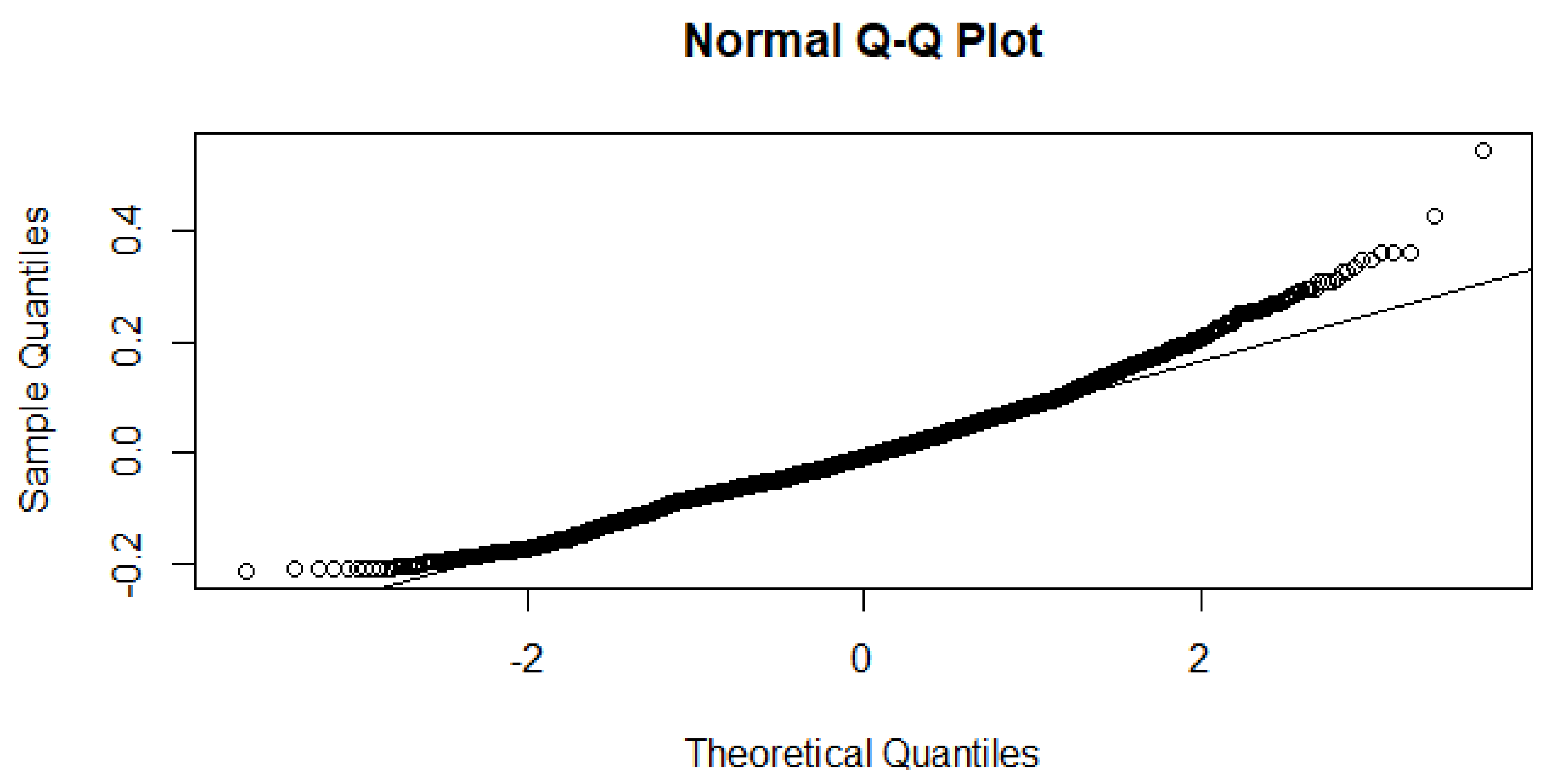
```
> model2  <- lm(VisceraWeight ~ Type, data = abalone)
> shapiro.test(residuals(model2))

    Shapiro-Wilk normality test

data:  residuals(model2)
W = 0.9825, p-value < 2.2e-16
```

03 Independence

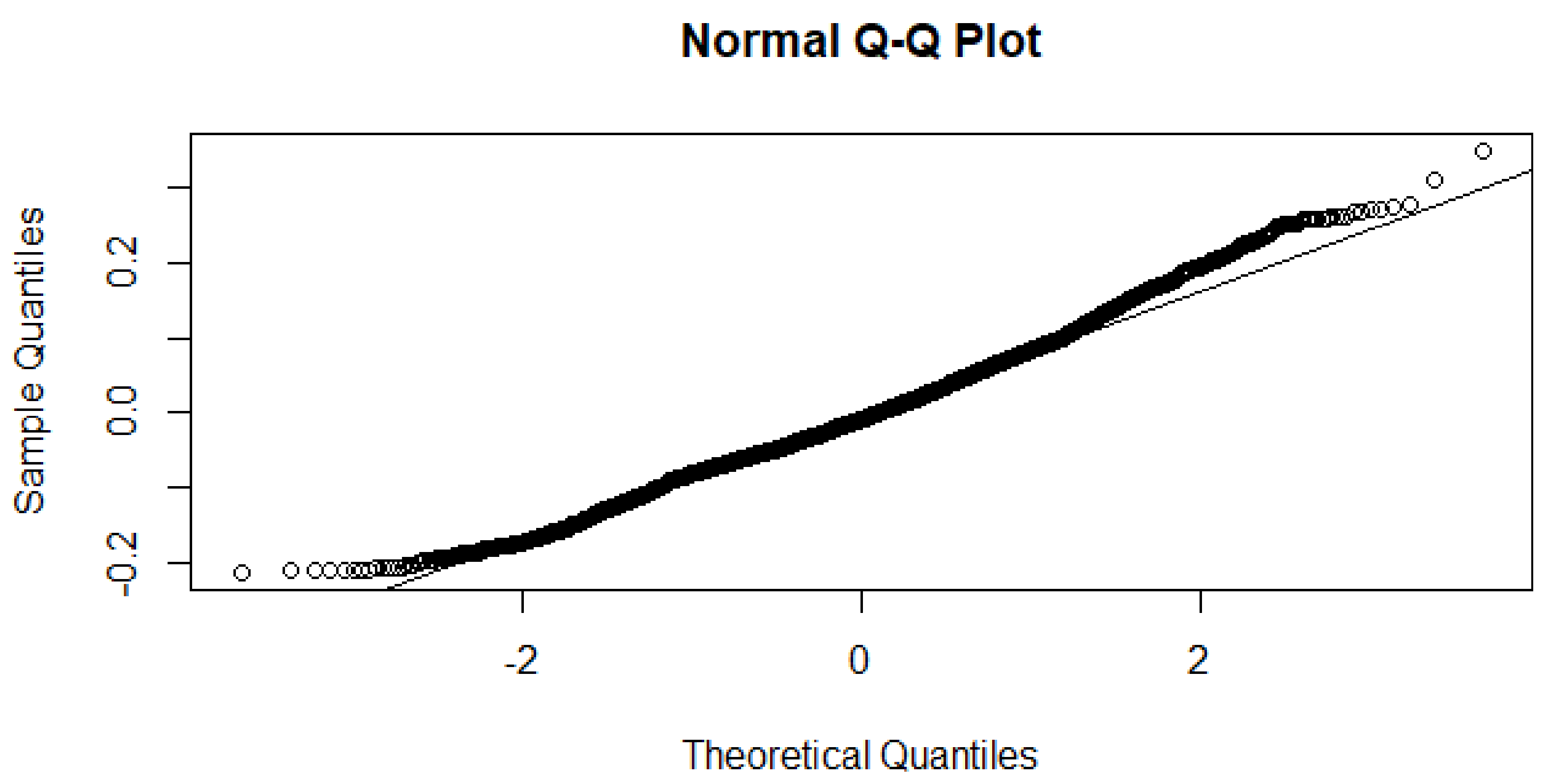
Viscera Weight

ANOVA

01 w/ outlier



02 w/o outlier



One-way ANOVA

Shucked Weight

```
> # Shucked
> # We will use Welch ANOVA.
> one_way2 <- oneway.test(ShuckedWeight~Type, data = data, var.equal = FALSE)
> one_way2

	One-way analysis of means (not assuming equal variances)

data:  ShuckedWeight and Type
F = 1108.3, num df = 2.0, denom df = 2646.8, p-value < 2.2e-16
```

Viscera Weight

```
> # Viscera
> # We will use Welch ANOVA.
> one_way3 <- oneway.test(VisceraWeight~Type, data = data, var.equal = FALSE)
> one_way3

	One-way analysis of means (not assuming equal variances)

data:  VisceraWeight and Type
F = 1321.3, num df = 2.0, denom df = 2649.9, p-value < 2.2e-16
```

Interval Estimate 03

1 **Shucked Weight**

ตัวแปรต้น : Sex (F,M,I)
ตัวแปรตาม : Shucked Weight

2 **Viscera Weight**

ตัวแปรต้น : Sex (F,M,I)
ตัวแปรตาม : Viscera Weight

เพศไหน มีน้ำหนักเฉลี่ยส่วนเนื้อหอยที่เยอะที่สุด

เพศไหน มีน้ำหนักเฉลี่ยส่วนอวัยวะภายในที่เยอะที่สุด

Shucked Weight (F,M)

```
 # Case I: F, M
> sd(abalone_F$ShuckedWeight) ; sd(abalone_M$ShuckedWeight) #sd(M) > sd(F)
[1] 0.1986632
[1] 0.223
> sd(abalone_M$ShuckedWeight)/sd(abalone_F$ShuckedWeight) #sample large-to-small SD ratio <2, it's
equal variances
[1] 1.122503
> t.test(x=abalone_F$ShuckedWeight,y=abalone_M$ShuckedWeight
+         ,alternative = 'two.sided',conf.level = 0.9,var.equal = T)

	Two Sample t-test

data:  abalone_F$ShuckedWeight and abalone_M$ShuckedWeight
t = 1.6568, df = 2833, p-value = 0.09767
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 9.126322e-05 2.639239e-02
sample estimates:
mean of x mean of y
0.4461878 0.4329460
```

Shucked Weight (F,I)

```
# Case II: F, I
> sd(abalone_F$ShuckedWeight) ; sd(abalone_I$ShuckedWeight) #sd(F) > sd(I)
[1] 0.1986632
[1] 0.1284053
> sd(abalone_F$ShuckedWeight)/sd(abalone_I$ShuckedWeight) #sample large-to-small SD ratio <2, it's
equal variances
[1] 1.547157
> t.test(x=abalone_F$ShuckedWeight,y=abalone_I$ShuckedWeight
+        ,alternative = 'two.sided',conf.level = 0.95,var.equal = T)

    Two Sample t-test

data:  abalone_F$ShuckedWeight and abalone_I$ShuckedWeight
t = 39.36, df = 2647, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2424413 0.2678643
sample estimates:
mean of x mean of y
0.4461878 0.1910350
```

Shucked Weight (M,I)

```
 # Case III: M, I
> sd(abalone_M$ShuckedWeight) ; sd(abalone_I$ShuckedWeight) #sd(M) > sd(I)
[1] 0.223
[1] 0.1284053
> sd(abalone_M$ShuckedWeight)/sd(abalone_I$ShuckedWeight) #sample large-to-small SD ratio <2, it's
equal variances
[1] 1.736688
> t.test(x=abalone_M$ShuckedWeight,y=abalone_I$ShuckedWeight
+        ,alternative = 'two.sided',conf.level = 0.95,var.equal = T)

    Two Sample t-test

data:  abalone_M$ShuckedWeight and abalone_I$ShuckedWeight
t = 34.972, df = 2868, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2283478 0.2554741
sample estimates:
mean of x mean of y
 0.432946  0.191035
```

03

Interval Estimate

- **Shucked Weight**

(Two Means t-test two-sided)

Sex	Hypothesis	P-value	Confidence interval
F,M	H0 : μF - μM = 0 H1 : μF - μM ≠ 0	p = 0.09767	90% CI: (0.00009126322, 0.02639239)
F,I	H0 : μF - μI = 0 H1 : μF - μI ≠ 0	p-value < 2.2e-16	90% CI: (0.2444862 , 0.2658195)
M,I	H0 : μM - μI = 0 H1 : μM - μI ≠ 0	p-value < 2.2e-16	90% CI: (0.2305296 0.2532924)

Viscera Weight (F,M)

```
> # Case I: F, M
> sd(abalone_F$VisceraWeight) ; sd(abalone_M$VisceraWeight) #sd(M) > sd(F)
[1] 0.09761738
[1] 0.104919
> sd(abalone_M$VisceraWeight)/sd(abalone_F$VisceraWeight) #sample large-to-small SD ratio <2, it's
equal variances
[1] 1.074798
> t.test(x=abalone_F$VisceraWeight,y=abalone_M$VisceraWeight
+        ,alternative = 'two.sided',conf.level = 0.95,var.equal = T)

    Two Sample t-test

data:  abalone_F$VisceraWeight and abalone_M$VisceraWeight
t = 3.9554, df = 2833, p-value = 7.827e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.007636825 0.022651369
sample estimates:
mean of x mean of y
0.2306886 0.2155445
```

Viscera Weight (F,I)

```
>> # Case II: F, I
> sd(abalone_F$VisceraWeight) ; sd(abalone_I$VisceraWeight) #sd(F) > sd(I)
[1] 0.09761738
[1] 0.06253609
> sd(abalone_F$VisceraWeight)/sd(abalone_I$VisceraWeight) #sample large-to-small SD ratio <2, it's
equal variances
[1] 1.560977
> t.test(x=abalone_F$VisceraWeight,y=abalone_I$VisceraWeight
+         ,alternative = 'two.sided',conf.level = 0.95,var.equal = T)

	Two Sample t-test

data:  abalone_F$VisceraWeight and abalone_I$VisceraWeight
t = 43.652, df = 2647, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.132449 0.144908
sample estimates:
 mean of x  mean of y
0.23068860 0.09201006
```

Viscera Weight (M,I)

```
>> # Case III: M, I
> sd(abalone_M$VisceraWeight) ; sd(abalone_I$VisceraWeight) #sd(M) > sd(I)
[1] 0.104919
[1] 0.06253609
> sd(abalone_M$VisceraWeight)/sd(abalone_I$VisceraWeight) #sample large-to-small SD ratio <2, it's
equal variances
[1] 1.677735
> t.test(x=abalone_M$VisceraWeight,y=abalone_I$VisceraWeight
+        ,alternative = 'two.sided',conf.level = 0.95,var.equal = T)

    Two Sample t-test

data:  abalone_M$VisceraWeight and abalone_I$VisceraWeight
t = 37.656, df = 2868, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1171019 0.1299670
sample estimates:
 mean of x  mean of y
0.21554450 0.09201006
```

02

Interval Estimate

- **Viscera Weight**

(Two Means t-test two-sided)

Sex	Hypothesis	P-value	Confidence interval
F,M	$H0: \mu F - \mu M = 0$ $H1: \mu F - \mu M \neq 0$	$p = 0.00007827$	95% CI: (0.007636825,0.022651369)
F,I	$H0: \mu F - \mu I = 0$ $H1: \mu F - \mu I \neq 0$	$p\text{-value} < 2.2e\text{-}16$	95% CI: (0.132449, 0.144908)
M,I	$H0: \mu M - \mu I = 0$ $H1: \mu M - \mu I \neq 0$	$p\text{-value} < 2.2e\text{-}16$	95% CI: (0.1171019, 0.1299670)

04

Multiple Linear Regression

- ## ตัวแปรสุ่มที่ต้องการศึกษา

Shucked Weight

ตัวแปรต้น : ความยาวของเปลือกหอย(Length), เส้นผ่านศูนย์กลางของเปลือกหอย(Diameter),ความสูงของเปลือกหอย(Height),น้ำหนักทั้งหมดของหอยเป๋าฮื้อ(Whole.weight),เพศ(Type)
ตัวแปรตาม : น้ำหนักเนื้อหอยเป๋าฮื้อ (Shucked.weight)

Viscera Weight

ตัวแปรต้น : ความยาวของเปลือกหอย(Length), เส้นผ่านศูนย์กลางของเปลือกหอย(Diameter),ความสูงของเปลือกหอย(Height),น้ำหนักทั้งหมดของหอยเป๋าฮื้อ(Whole.weight),เพศ(Type)
ตัวแปรตาม : น้ำหนักอวัยวะภายในเนื้อหอยเป๋าฮื้อ (Viscera.weight)

Multiple Linear Regression

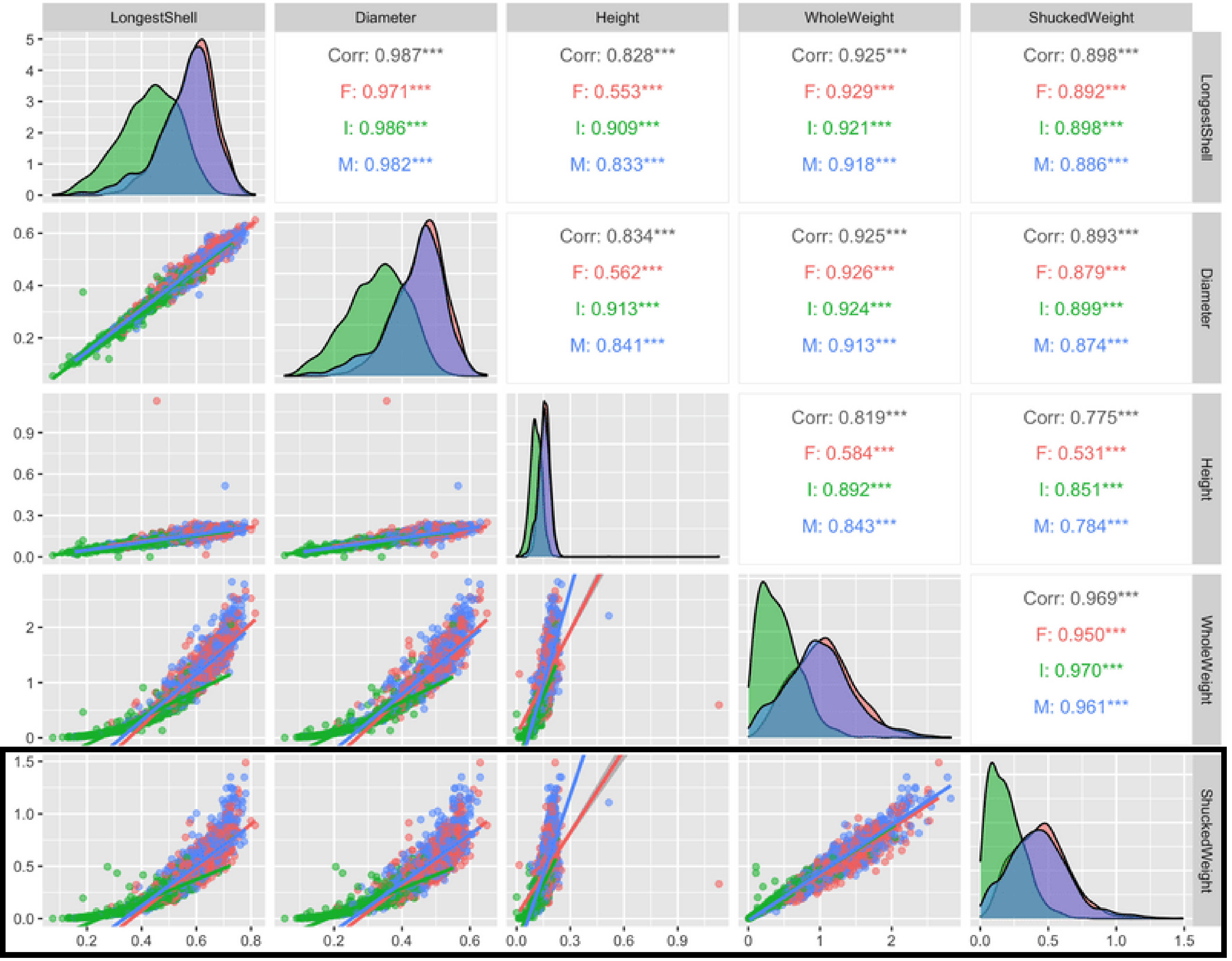
- **สมมติฐาน**

H0 : ตัวแปรอิสระทุกตัวไม่สามารถร่วมกันทำนายพยากรณ์ตัวแปรตามได้ (No effect)

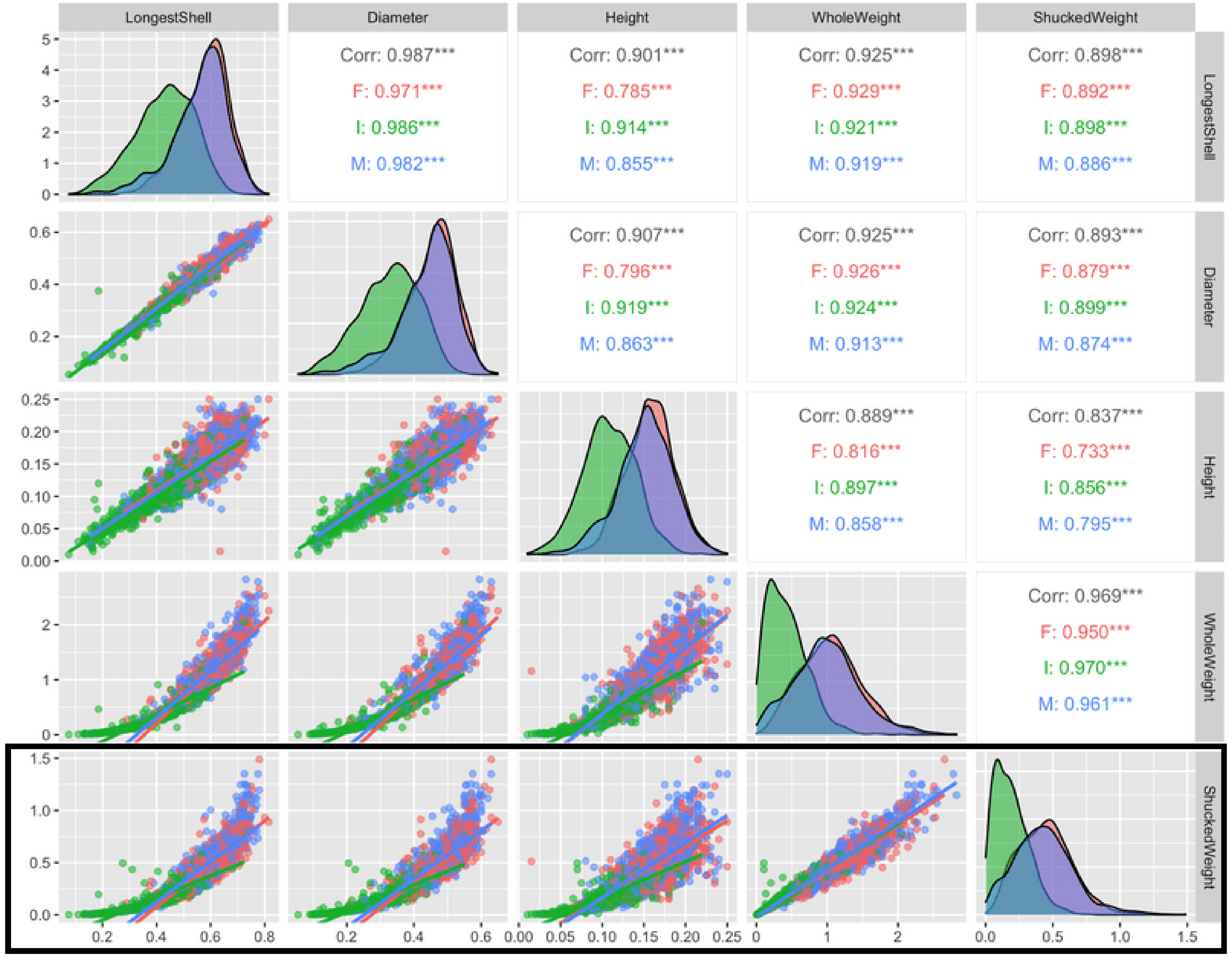
H1 : ตัวแปรอิสระทุกตัวสามารถร่วมกันทำนายพยากรณ์ตัวแปรตามได้ (At the least one ≠ 0)

Multiple Linear Regression

- ## Shucked Weight



Plot with outlier



Plot without outlier

Multiple Linear Regression

- **Shucked Weight**

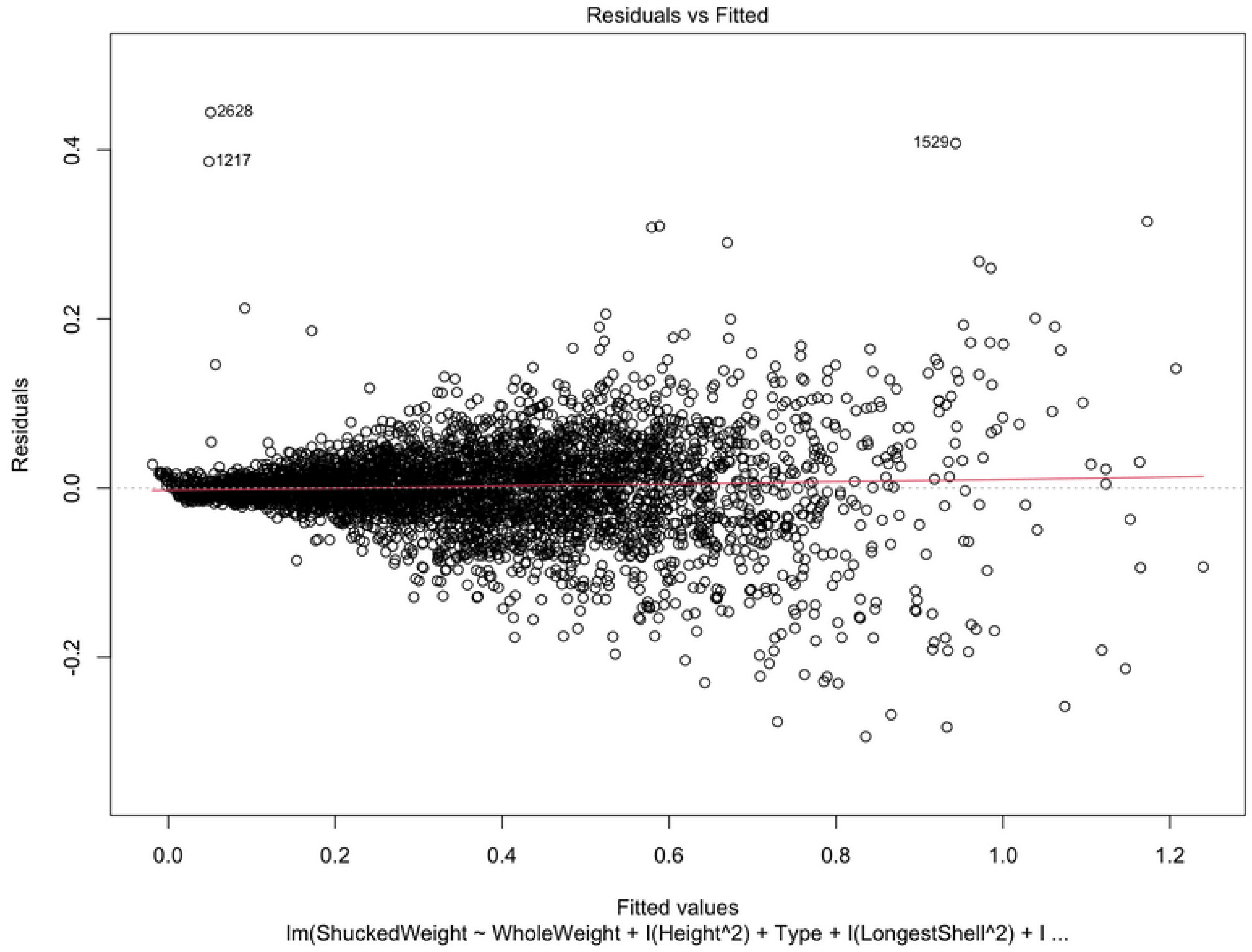
Stepwise AIC selection(normal)

```
Console   Terminal ×   Background Jobs ×
R 4.2.1 · ~/
> summary(abalone.step)

Call:
lm(formula = ShuckedWeight ~ WholeWeight + I(Height^2) + Type +
    I(LongestShell^2) + I(Diameter^2) + LongestShell + Height +
    I(Height^3) + Type:LongestShell, data = abalone0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.29403 -0.01949 -0.00086  0.02237  0.44423

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.003491   0.022775   0.153 0.878190
WholeWeight         0.467627   0.007451  62.764  < 2e-16 ***
I(Height^2)       -10.246723   3.399924  -3.014 0.002595 **
TypeI               0.003318   0.014260   0.233 0.816012
TypeM              -0.031144   0.012626  -2.467 0.013678 *
I(LongestShell^2)   0.580848   0.097355   5.966 2.63e-09 ***
I(Diameter^2)      -0.295826   0.061820  -4.785 1.77e-06 ***
LongestShell       -0.338090   0.097704  -3.460 0.000545 ***
Height              1.396760   0.498495   2.802 0.005103 **
I(Height^3)        12.866445   7.754984   1.659 0.097167 .
TypeI:LongestShell  0.015003   0.027383   0.548 0.583784
TypeM:LongestShell  0.071965   0.021749   3.309 0.000945 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05204 on 4161 degrees of freedom
Multiple R-squared:  0.945,     Adjusted R-squared:  0.9449
F-statistic:  6505 on 11 and 4161 DF,  p-value: < 2.2e-16

>
```

Multiple Linear Regression

- **Shucked Weight** : Residual Diagnostic

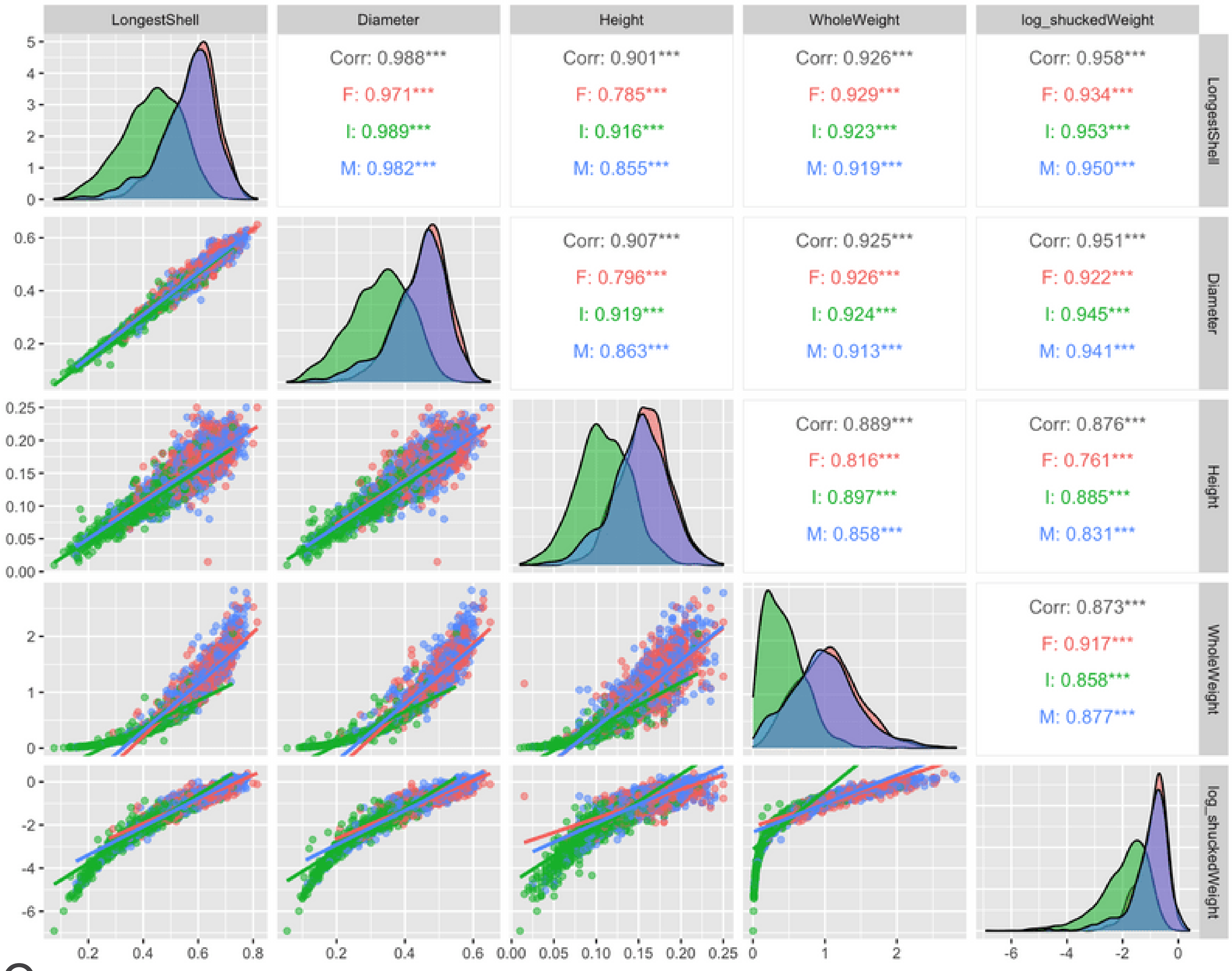
plot(abalone.step,which=1)



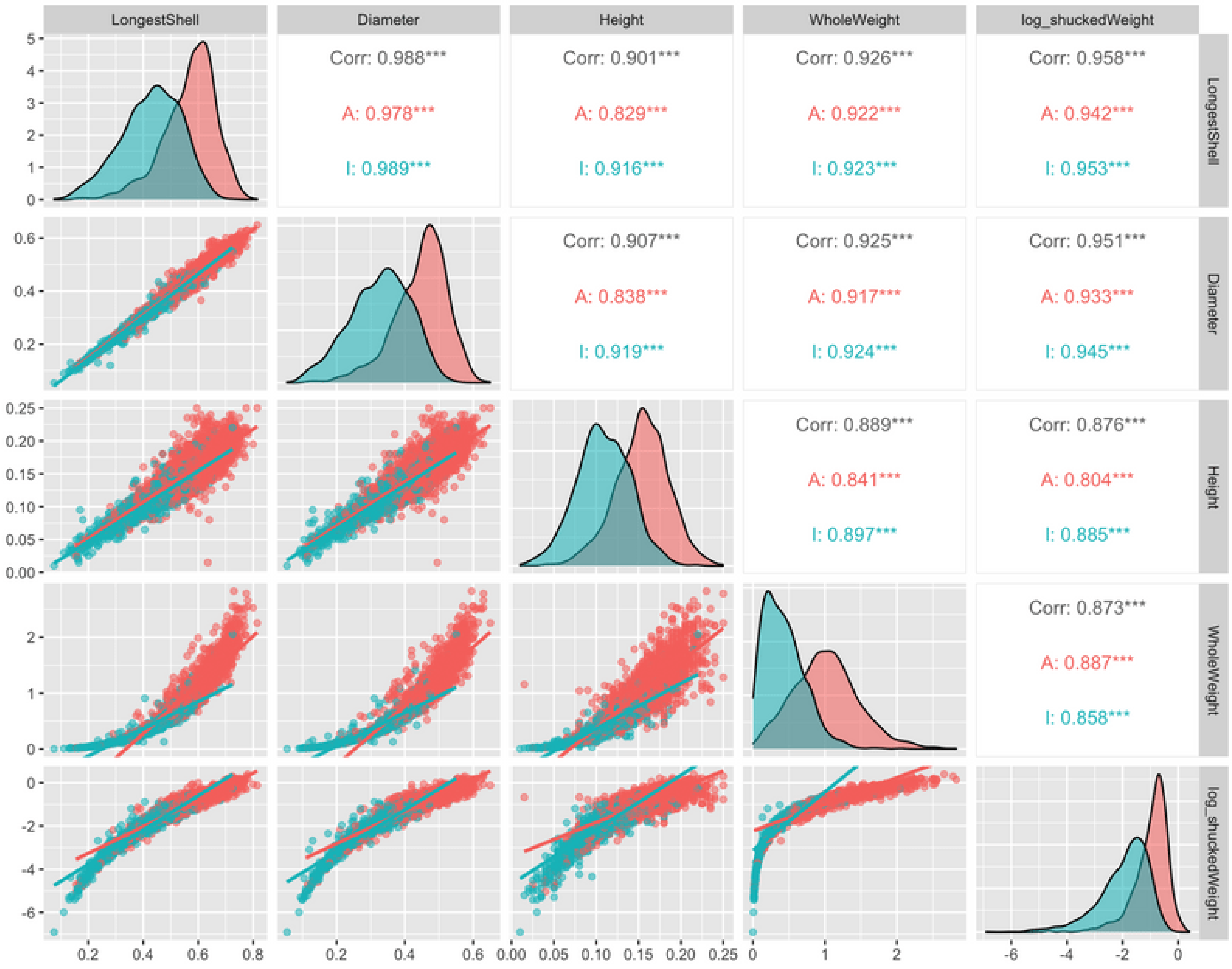
Multiple Linear Regression

- ## Shucked Weight

Predict log (Type F,M,I)



Predict log (Type F+M,I)



Multiple Linear Regression

- Shucked Weight

```r
abalone.null <- lm(log_shuckedWeight~1,data=abalone0)
abalone.step <- step(abalone.null,scope = .~.+LongestShell+I(LongestShell^2)+Type+
                                              Diameter+I(Diameter^2)+
                                              Height+I(Height^2)+WholeWeight+I(WholeWeight^2)+I(Height^3)+
                                              Type*LongestShell+Type*Diameter+Type*Height+Type*WholeWeight)

summary(abalone.step
```

Multiple Linear Regression

log_tran

- **Shucked Weight**

Stepwise AIC selection(normal)

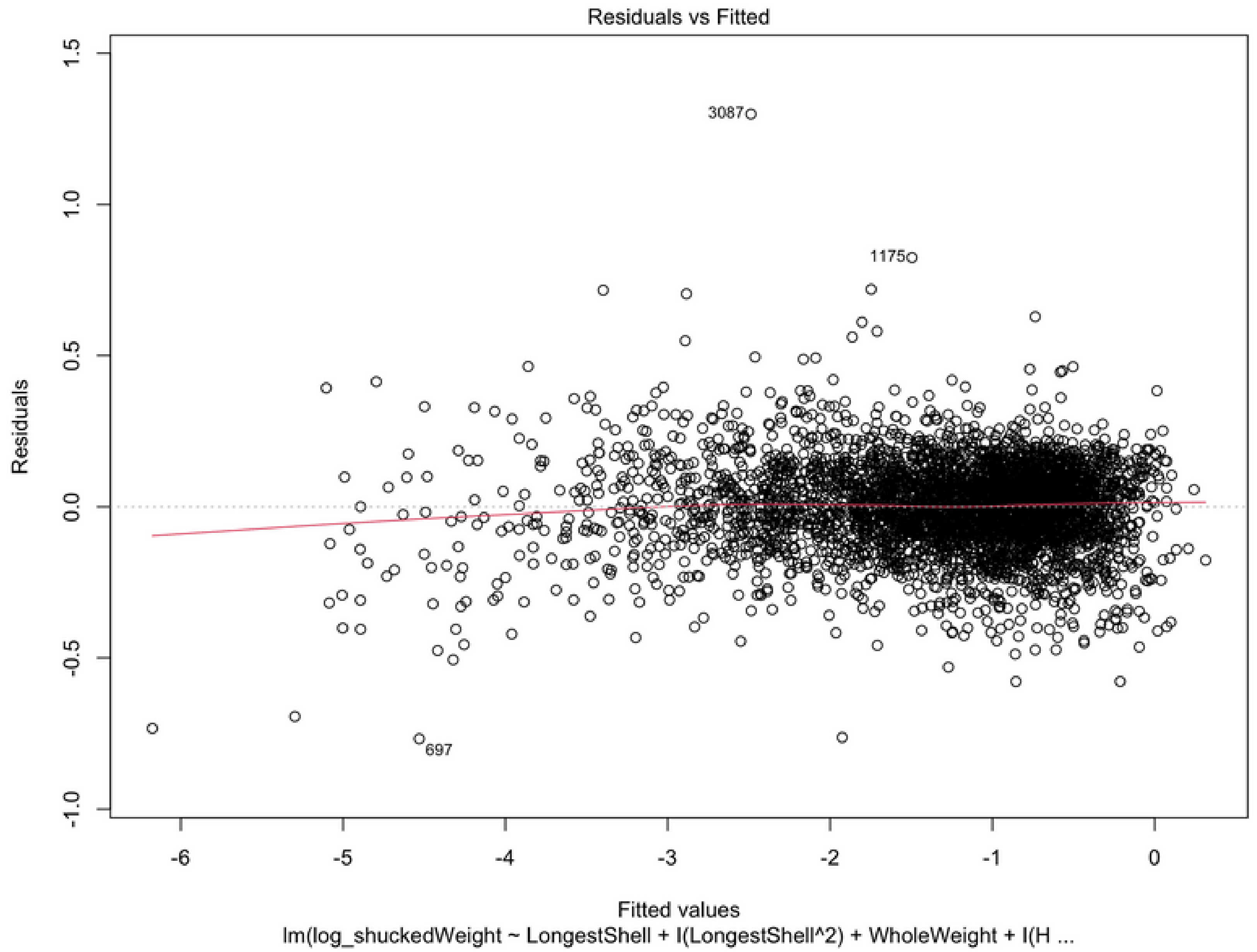
```
Console   Terminal ×   Background Jobs ×
R 4.2.1 · ~/
> summary(abalone.step)

Call:
lm(formula = log_shuckedWeight ~ LongestShell + I(LongestShell^2) +
    WholeWeight + I(Height^3) + I(WholeWeight^2) + Type + I(Diameter^2) +
    Diameter + Height + I(Height^2) + LongestShell:Type + WholeWeight:Type,
    data = abalone0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.76755 -0.08127  0.00789  0.08642  1.29858

Coefficients:
                    Estimate Std. Error  t value Pr(>|t|)
(Intercept)         -7.69263    0.06658 -115.542  < 2e-16 ***
LongestShell        12.51480    0.67821   18.453  < 2e-16 ***
I(LongestShell^2)   -9.97992    0.59016  -16.910  < 2e-16 ***
WholeWeight          1.58062    0.05184   30.490  < 2e-16 ***
I(Height^3)        472.79509   26.91166   17.568  < 2e-16 ***
I(WholeWeight^2)    -0.20069    0.01673  -11.993  < 2e-16 ***
TypeI                0.28792    0.04929    5.841 5.58e-09 ***
I(Diameter^2)       -3.03259    0.85207   -3.559 0.000376 ***
Diameter             2.09641    0.75400    2.780 0.005454 **
Height              31.29584    1.59894   19.573  < 2e-16 ***
I(Height^2)       -223.05682   11.49514  -19.404  < 2e-16 ***
LongestShell:TypeI  -0.70524    0.15285   -4.614 4.07e-06 ***
WholeWeight:TypeI    0.15377    0.04740    3.244 0.001187 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1436 on 4157 degrees of freedom
Multiple R-squared:  0.9723,    Adjusted R-squared:  0.9722
F-statistic: 1.215e+04 on 12 and 4157 DF,  p-value: < 2.2e-16

>
```

Multiple Linear Regression

log_tran

- **Shucked Weight** : Residual Diagnostic

plot(abalone.step,which=1)

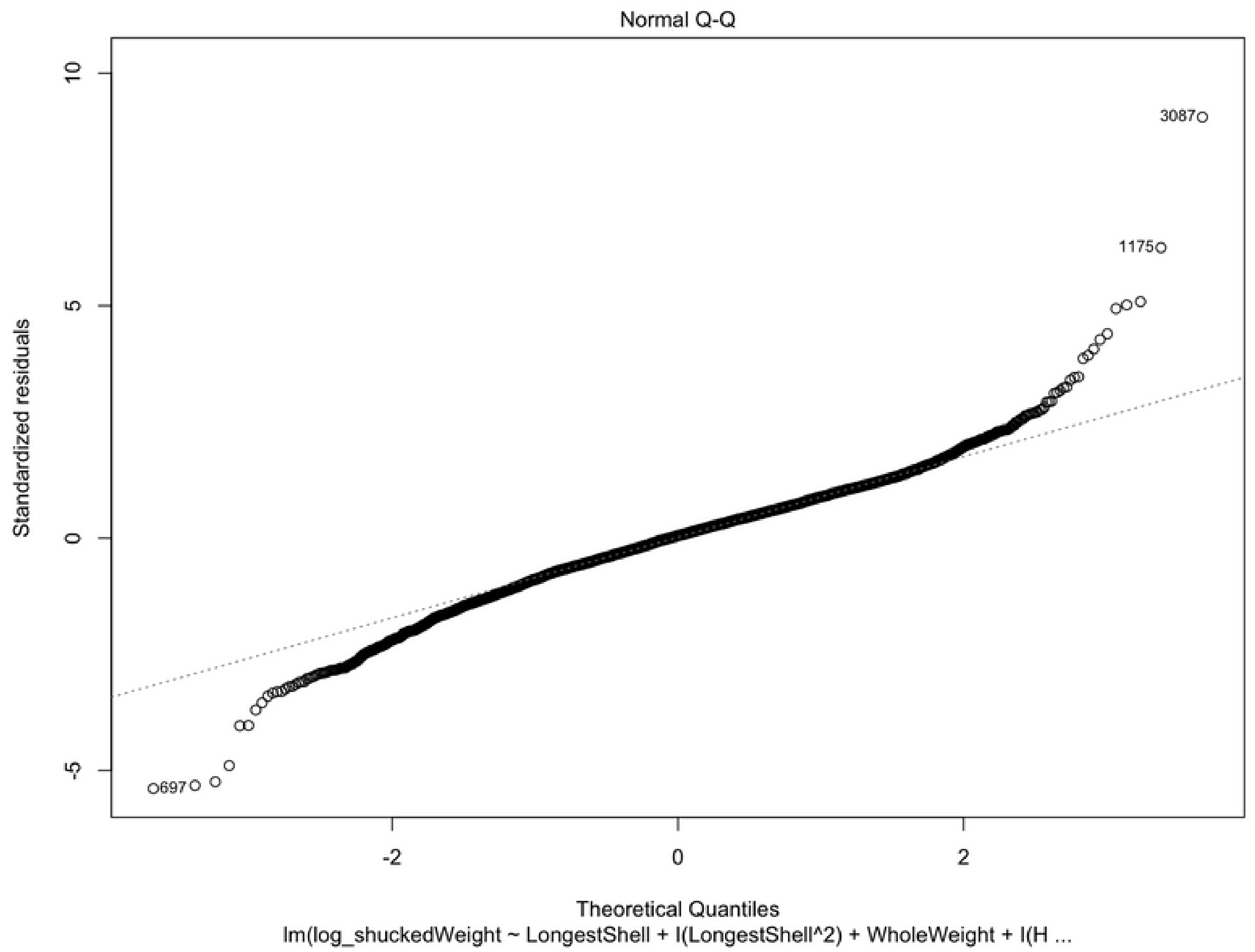


Multiple Linear Regression

log_tran

- **Shucked Weight** : Residual Diagnostic

plot(abalone.step,which=2)

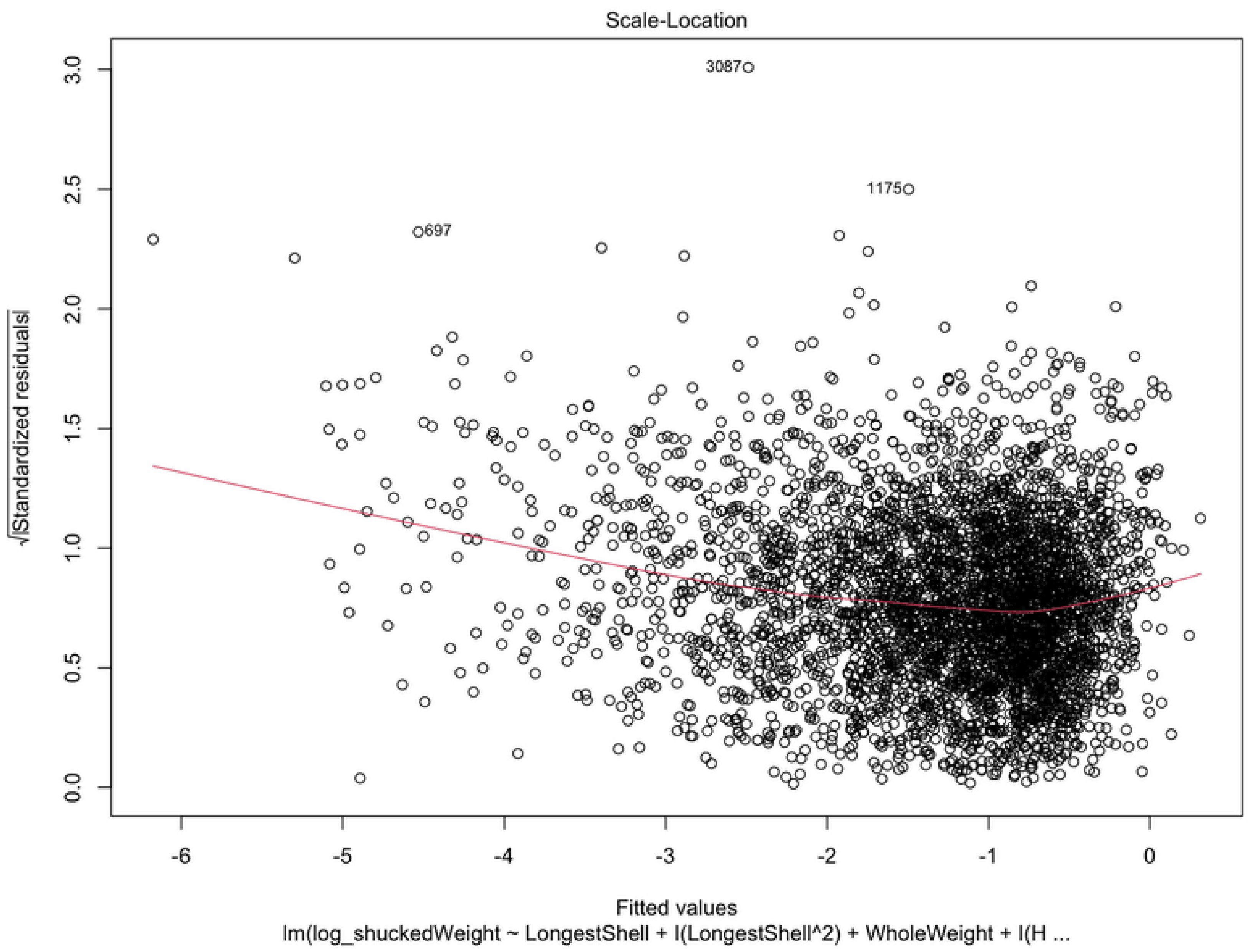


Multiple Linear Regression

log_tran

- **Shucked Weight** : Residual Diagnostic

plot(abalone.step,which=3)

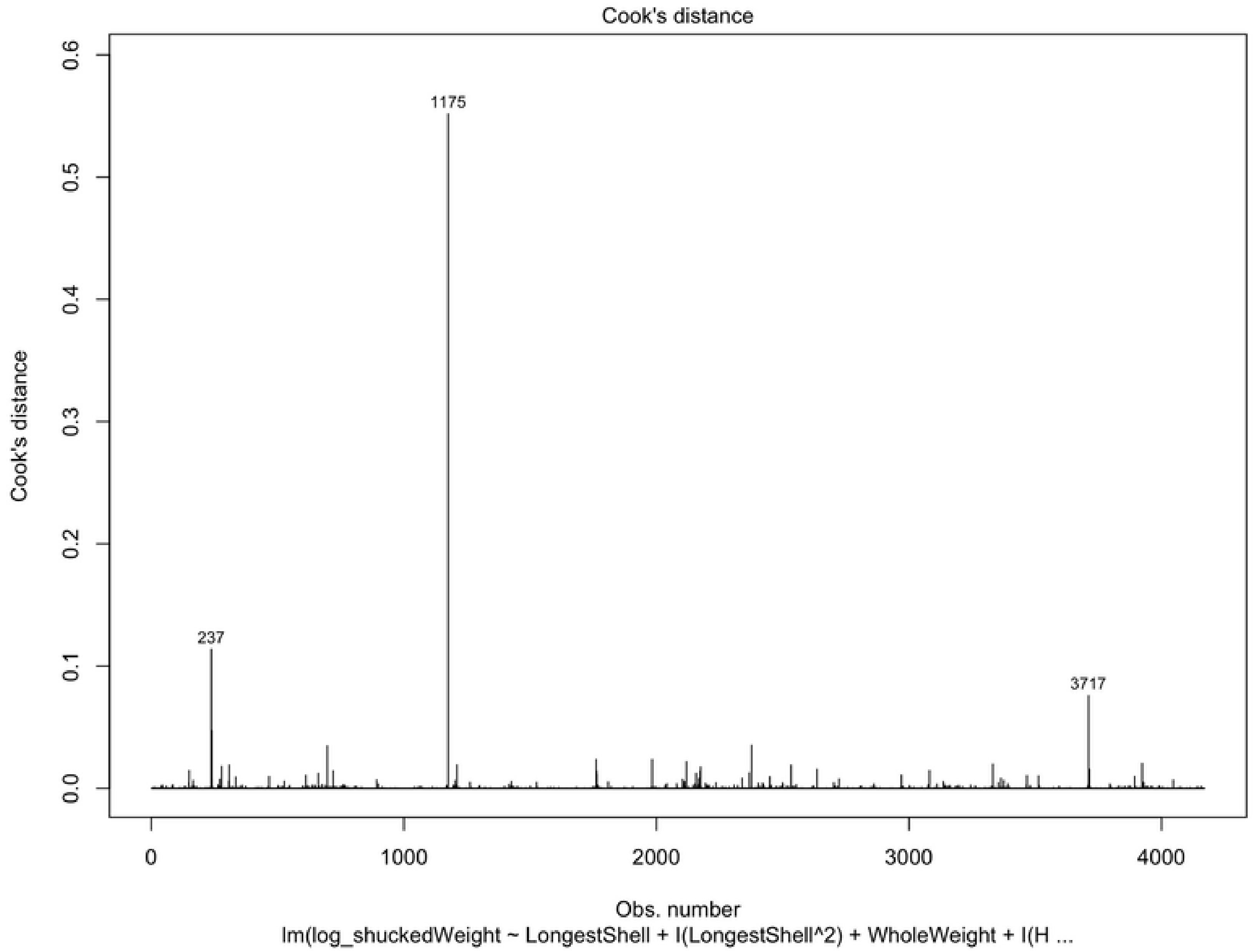


Multiple Linear Regression

log_tran

- **Shucked Weight** : Residual Diagnostic

plot(abalone.step,which=4)

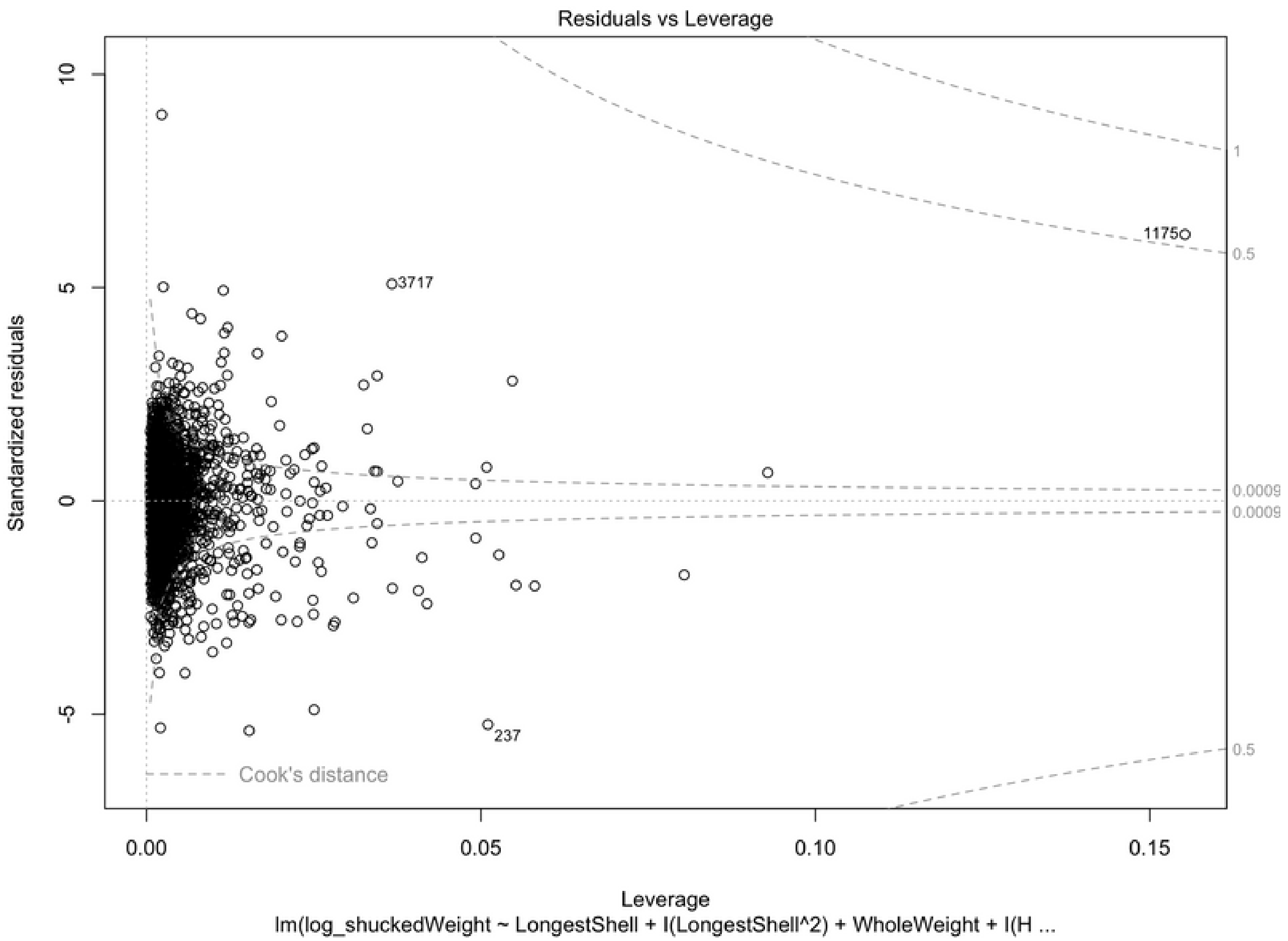


Multiple Linear Regression

log_tran

- **Shucked Weight** : Residual Diagnostic

plot(abalone.step,which=5)



Multiple Linear Regression

- Shucked Weight

General equation for fitted model is :

ln(ShuckedWeight) =

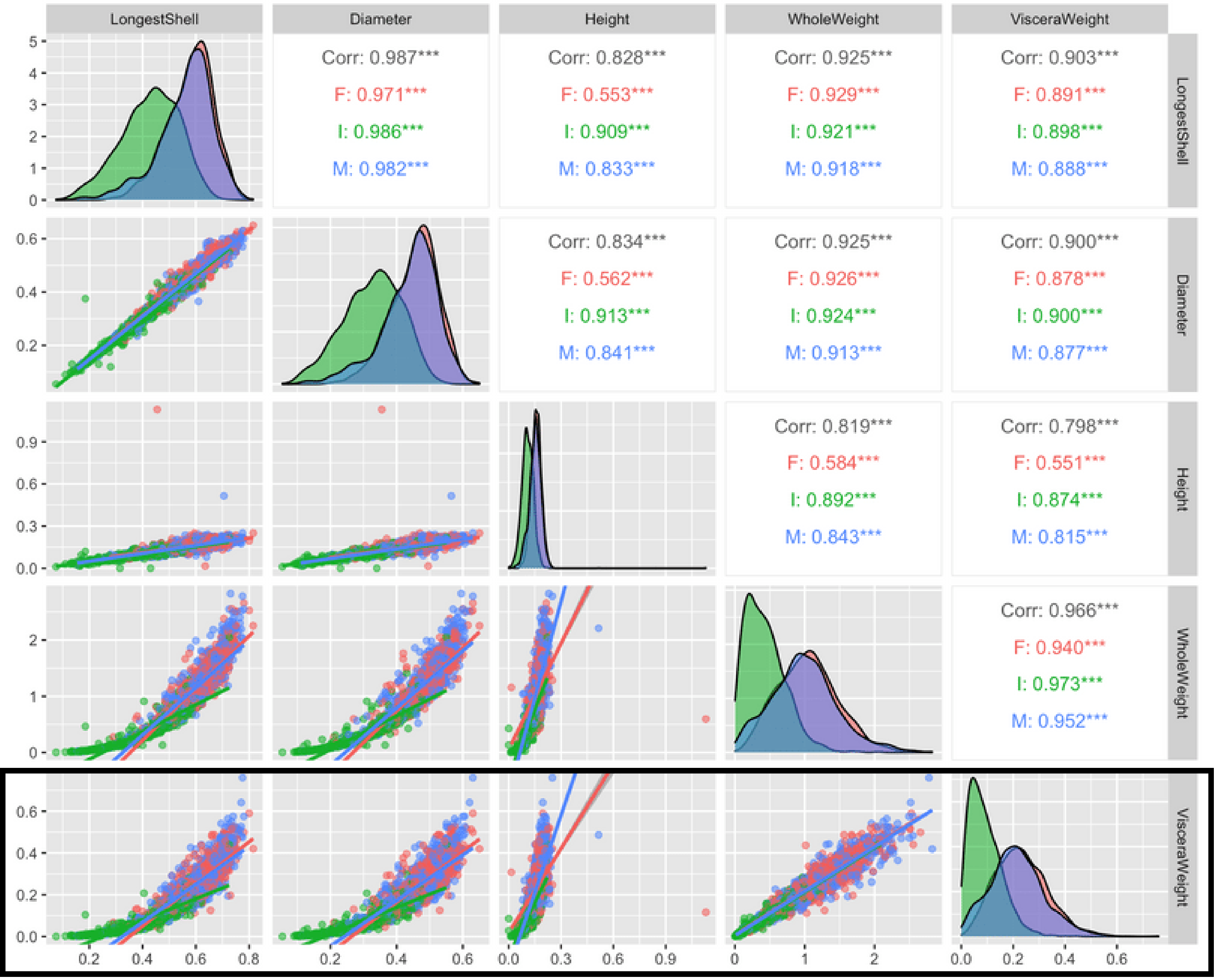
-7.69263 + 12.51480(LongestShell) - 9.97992(LongestShell^2)
+ 1.58062(WholeWeight) - 0.20069(WholeWeight^2)
+ 31.29584(Height) - 223.05682(Height^2) + 472.79509(Height^3)
+ 2.09641(Diameter) - 3.03259(Diameter^2) + 0.28792(TypeI)
- 0.70524(LongestShell of TypeI) + 0.15377(WholeWeight of Type I)

from, ln(x)=y ⟶ x=e^y

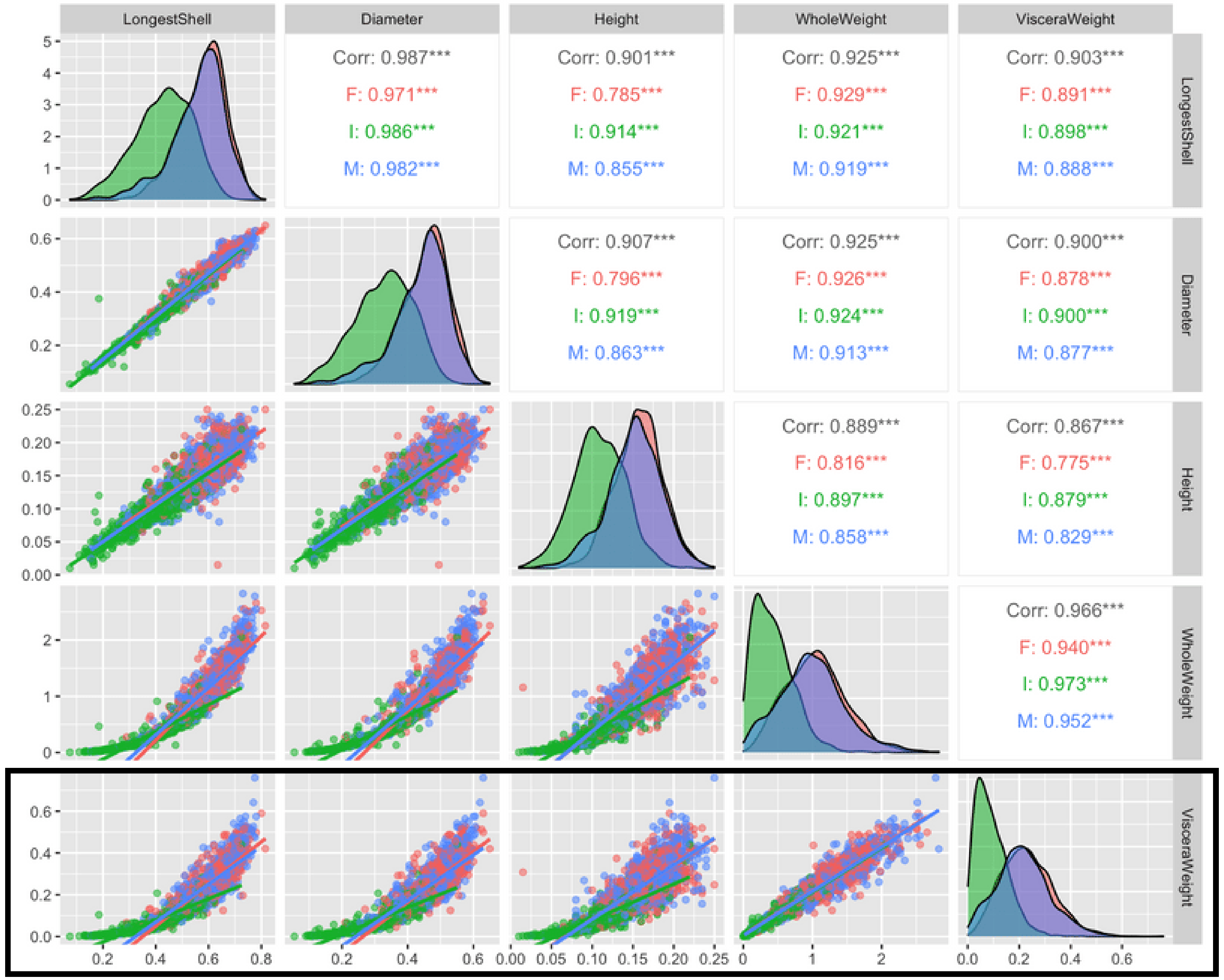
ShuckedWeight = e ^

Multiple Linear Regression

- ## Viscera Weight



Plot with outlier



Plot without outlier

Multiple Linear Regression

- **Viscera Weight**

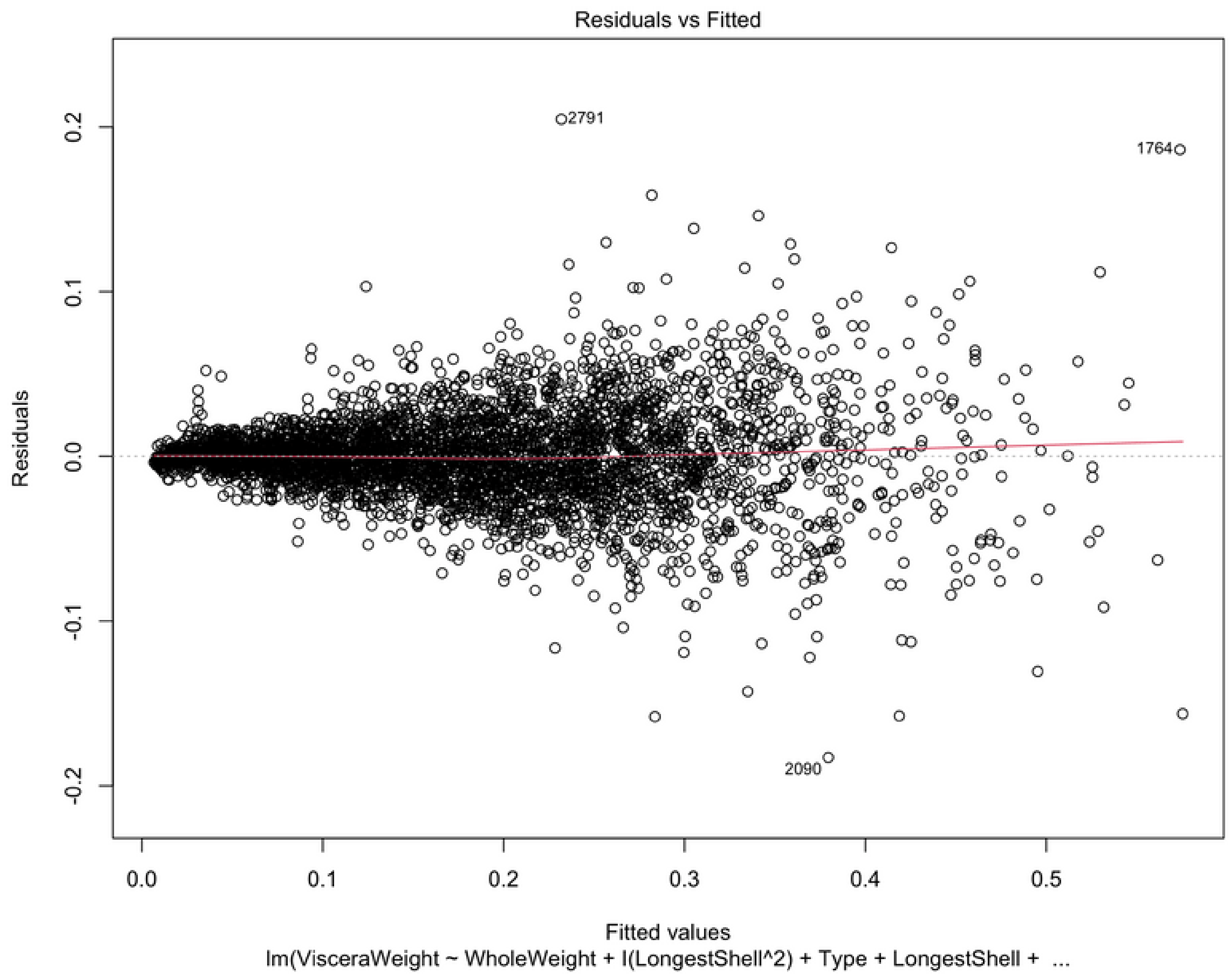
Stepwise AIC selection(normal)

```
Console   Terminal ×   Background Jobs ×
R 4.2.1 · ~/
> summary(abalone.step)

Call:
lm(formula = VisceraWeight ~ WholeWeight + I(LongestShell^2) +
    Type + LongestShell + I(Diameter^2) + I(Height^3) + I(Height^2) +
    Type:LongestShell, data = abalone0)

Residuals:
      Min        1Q    Median        3Q       Max
-0.182919 -0.013077 -0.000226  0.012044  0.204720

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.018914   0.011030   1.715 0.086451 .
WholeWeight         0.181852   0.003953  46.007  < 2e-16 ***
I(LongestShell^2)   0.343664   0.043851   7.837 5.82e-15 ***
TypeI               0.006678   0.007575   0.882 0.378084
TypeM               0.001109   0.006716   0.165 0.868846
LongestShell       -0.149378   0.042551  -3.511 0.000452 ***
I(Diameter^2)      -0.150300   0.032841  -4.577 4.86e-06 ***
I(Height^3)        -6.067914   1.912804  -3.172 0.001524 **
I(Height^2)         1.878304   0.485123   3.872 0.000110 ***
TypeI:LongestShell -0.028271   0.014534  -1.945 0.051830 .
TypeM:LongestShell -0.006860   0.011568  -0.593 0.553225
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02768 on 4162 degrees of freedom
Multiple R-squared:  0.9363,    Adjusted R-squared:  0.9361
F-statistic:  6115 on 10 and 4162 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

- **Viscera Weight** : Residual Diagnostic

plot(abalone.step,which=1)

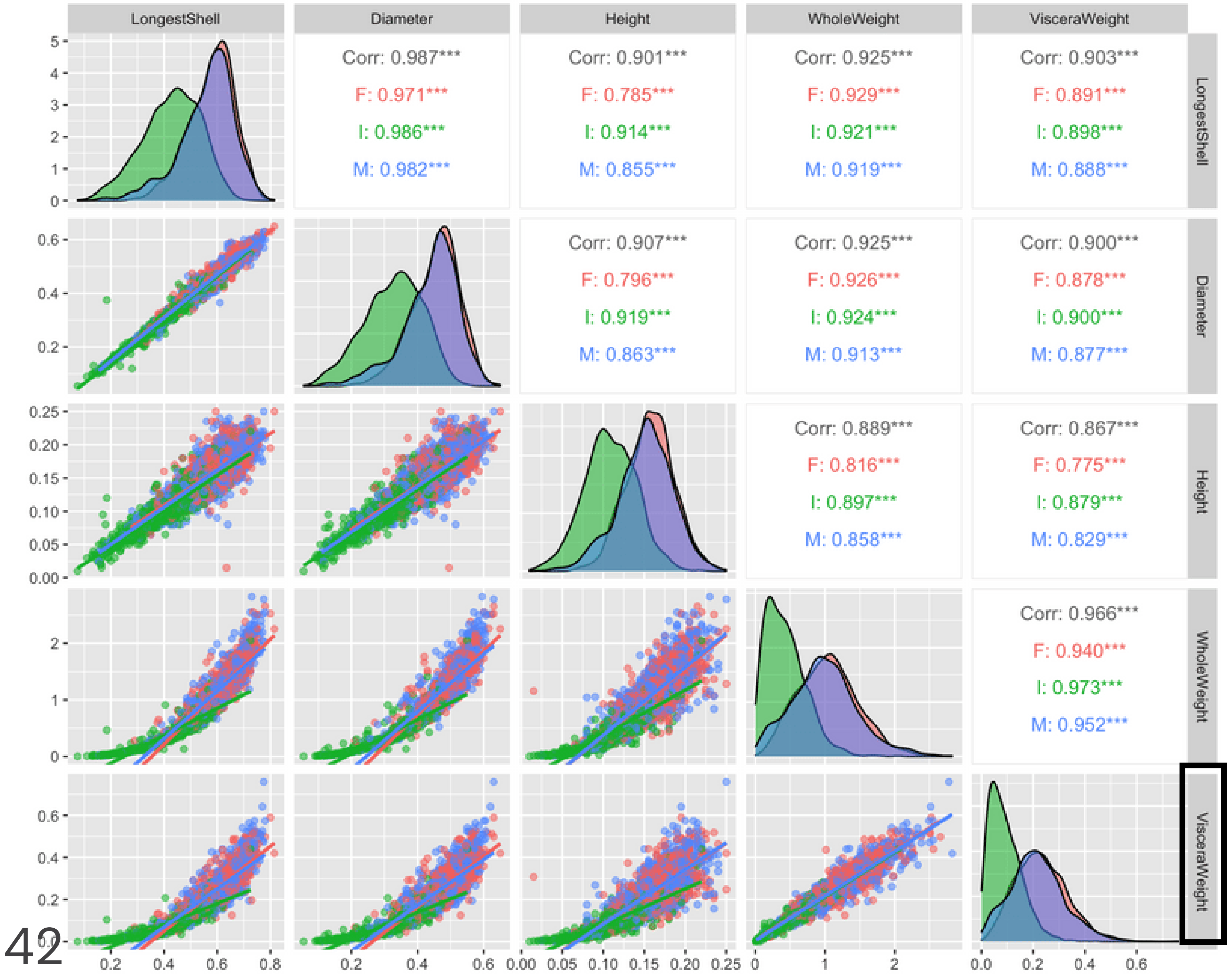


Heteroscedastic

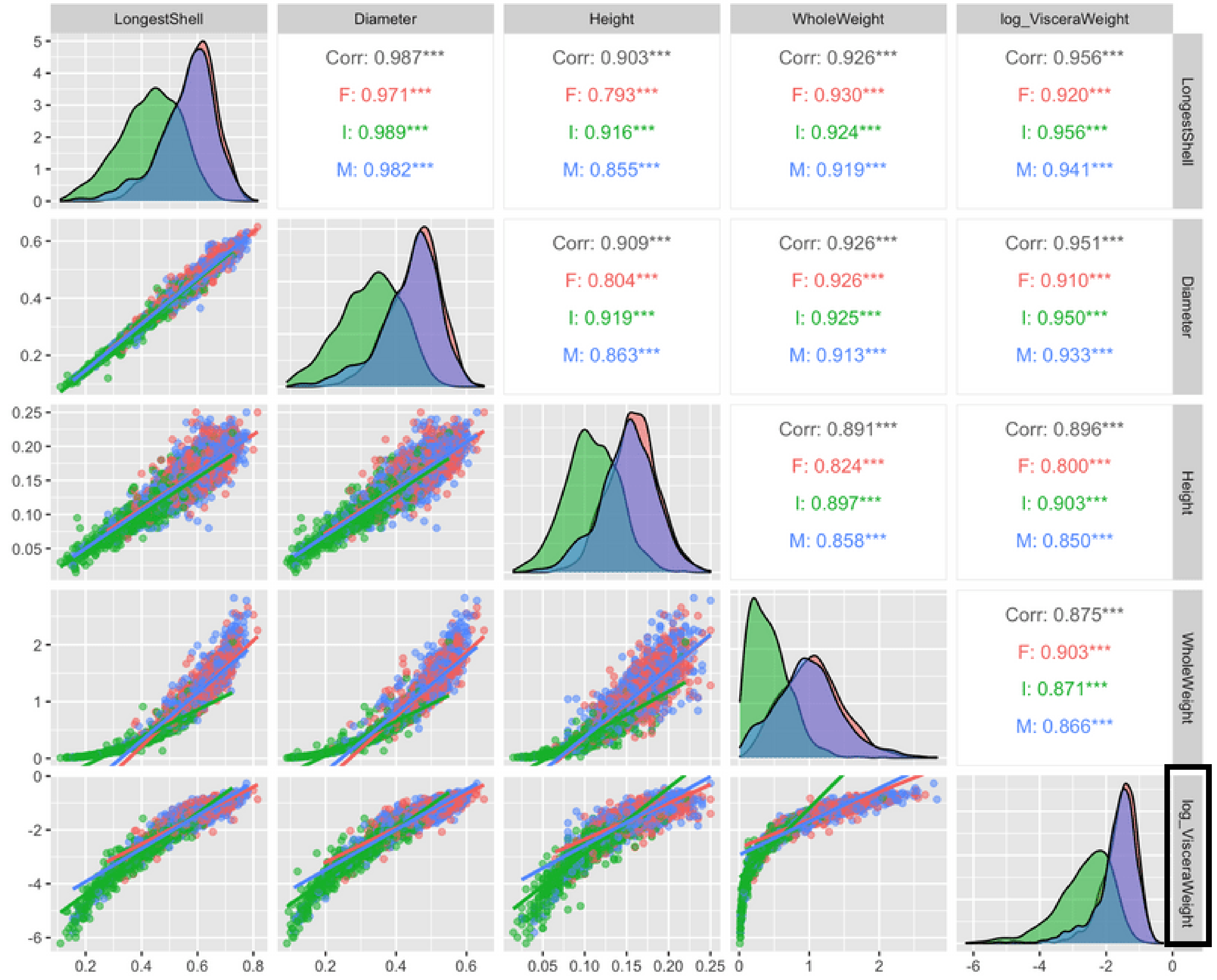
Multiple Linear Regression

- **Viscera Weight**

Initial version



Predict log version



Multiple Linear Regression

log_tran

- Viscera Weight

```
abalone.null <- lm(log_VisceraWeight~1,data=abalone0)
abalone.step <- step(abalone.null,scope = .~.+LongestShell+I(LongestShell^2)+Type+
                    Diameter+I(Diameter^2)+
                    Height+I(Height^2)+I(Height^3)+WholeWeight+I(WholeWeight^2)+

Type*LongestShell+Type*Diameter+Type*Height+Type*WholeWeight)
summary(abalone.step)
```

Multiple Linear Regression

log_tran

04

- **Viscera Weight**

Stepwise AIC selection(normal)

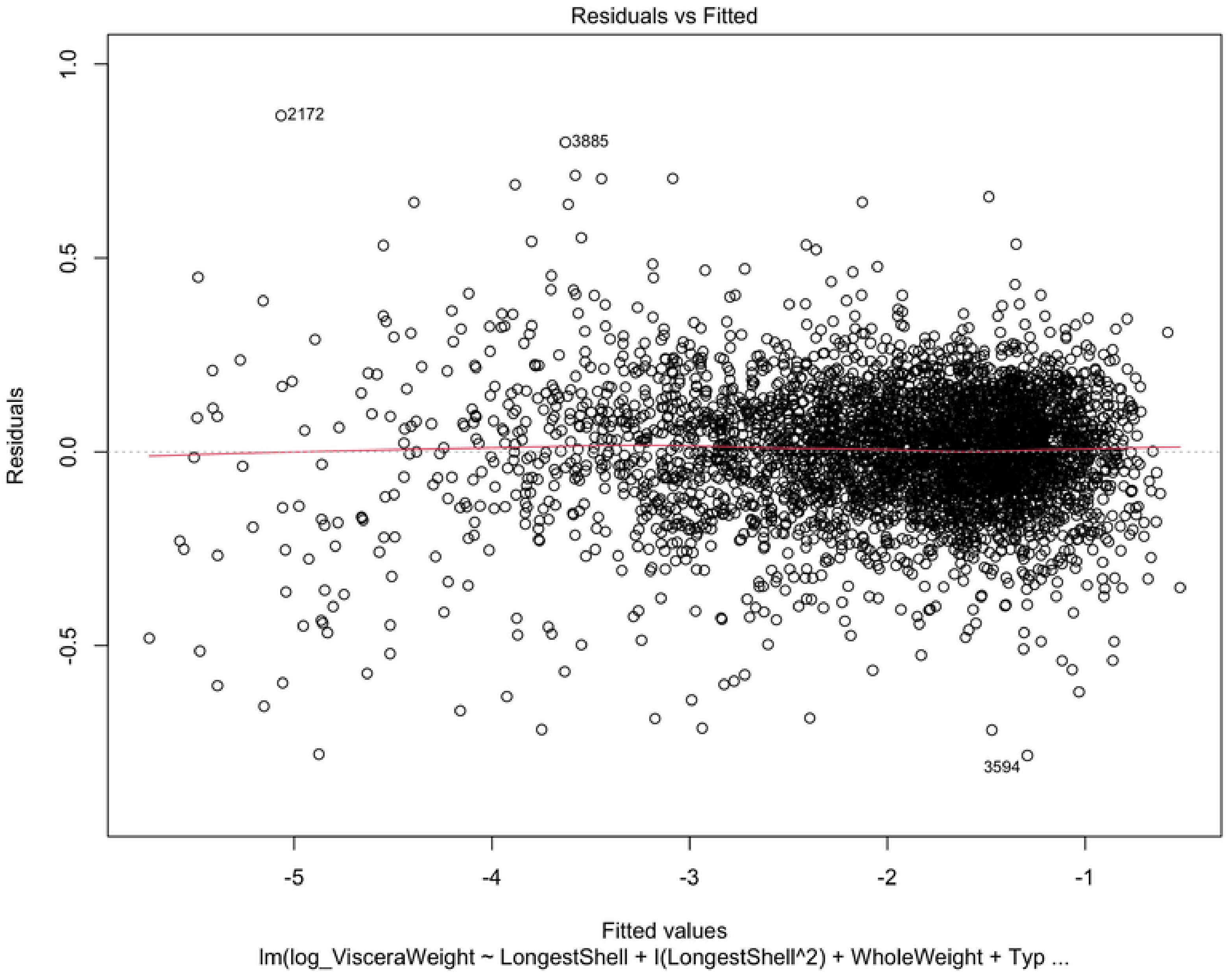
```
Console   Terminal ×   Background Jobs ×
R 4.2.1 · ~/
> summary(abalone.step)

Call:
lm(formula = log_VisceraWeight ~ LongestShell + I(LongestShell^2) +
    WholeWeight + Type + I(WholeWeight^2) + Height + I(Height^2) +
    I(Height^3) + I(Diameter^2) + Diameter, data = abalone0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.78350 -0.08985  0.00549  0.09701  0.86731

Coefficients:
                    Estimate Std. Error  t value Pr(>|t|)
(Intercept)       -7.499e+00  4.886e-02 -153.468  < 2e-16 ***
LongestShell       9.002e+00  7.244e-01   12.426  < 2e-16 ***
I(LongestShell^2) -6.966e+00  6.391e-01  -10.900  < 2e-16 ***
WholeWeight        1.302e+00  5.159e-02   25.230  < 2e-16 ***
TypeI             -5.554e-02  7.588e-03   -7.319 2.98e-13 ***
TypeM             -1.641e-02  6.009e-03   -2.731 0.006337 **
I(WholeWeight^2)  -1.751e-01  1.702e-02  -10.290  < 2e-16 ***
Height             2.665e+01  1.909e+00   13.955  < 2e-16 ***
I(Height^2)       -1.607e+02  1.347e+01  -11.928  < 2e-16 ***
I(Height^3)        3.169e+02  3.108e+01   10.197  < 2e-16 ***
I(Diameter^2)     -4.062e+00  9.500e-01   -4.276 1.95e-05 ***
Diameter           3.042e+00  8.419e-01    3.613 0.000307 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1585 on 4154 degrees of freedom
Multiple R-squared:  0.9645,    Adjusted R-squared:  0.9644
F-statistic: 1.027e+04 on 11 and 4154 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

log_tran

• Viscera Weight : Residual Diagnostic

plot(abalone.step,which=1)



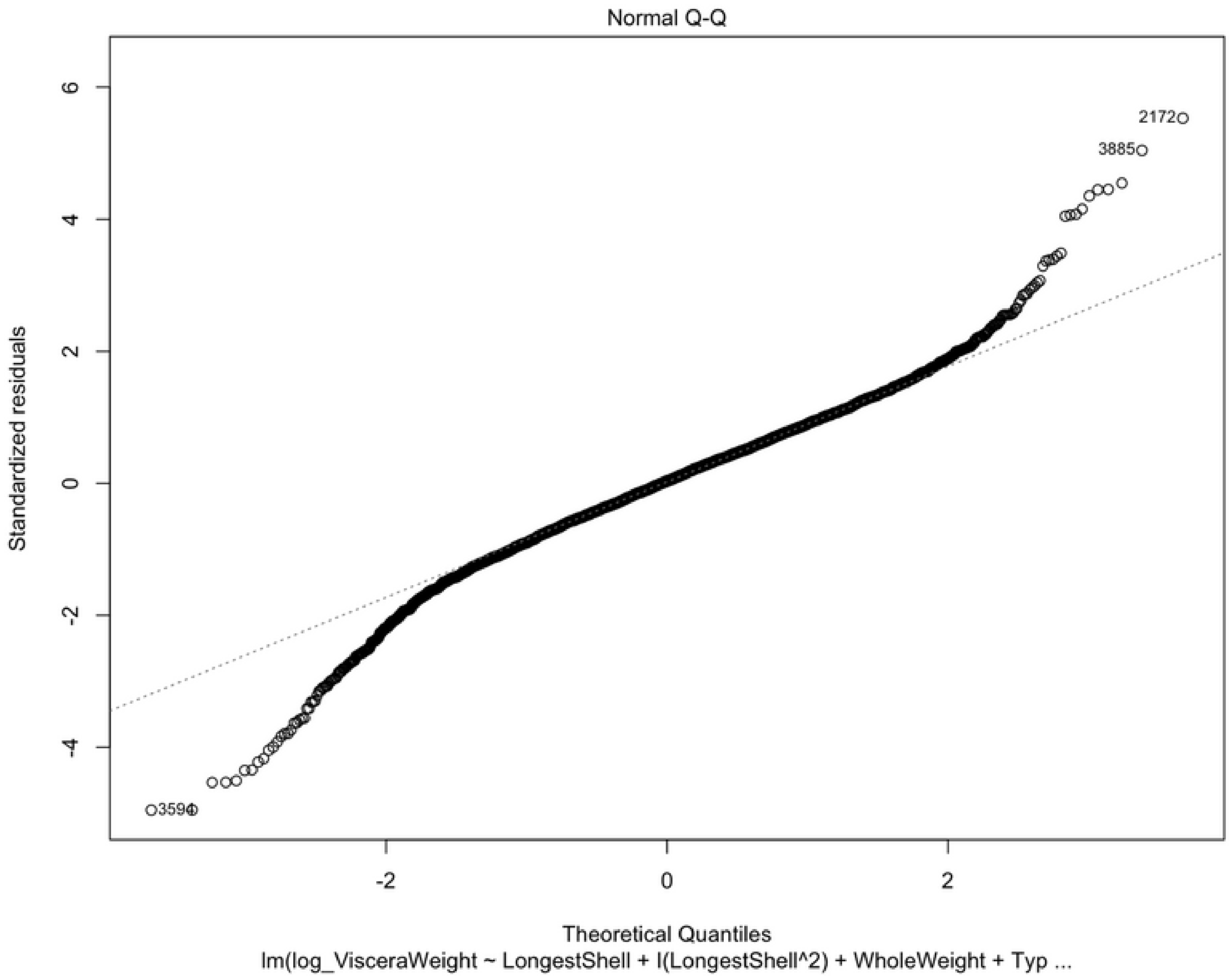
Multiple Linear Regression

60

log_tran

• Viscera Weight : Residual Diagnostic

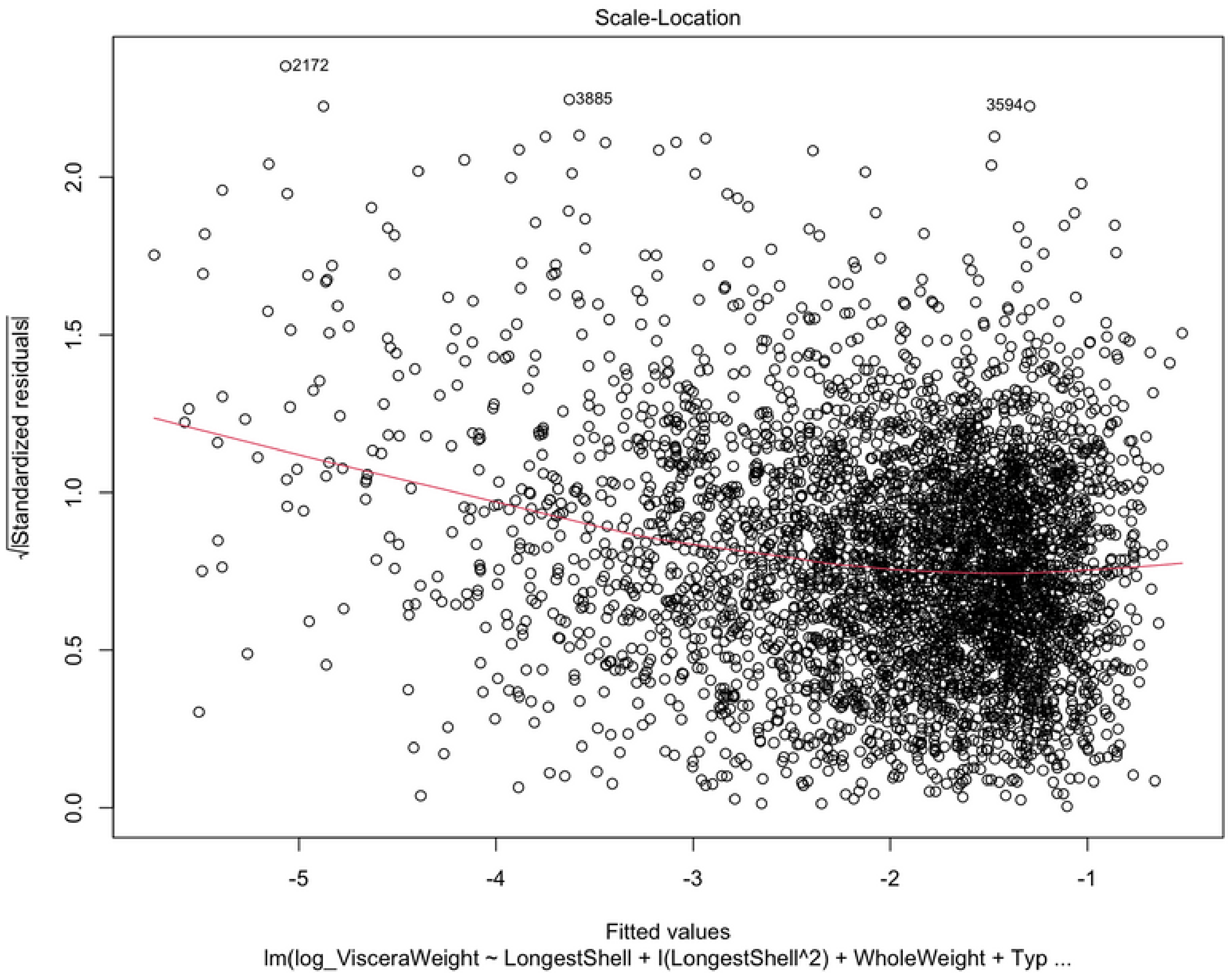
plot(abalone.step,which=2)



Multiple Linear Regression

- **Viscera Weight** : Residual Diagnostic

plot(abalone.step,which=3)

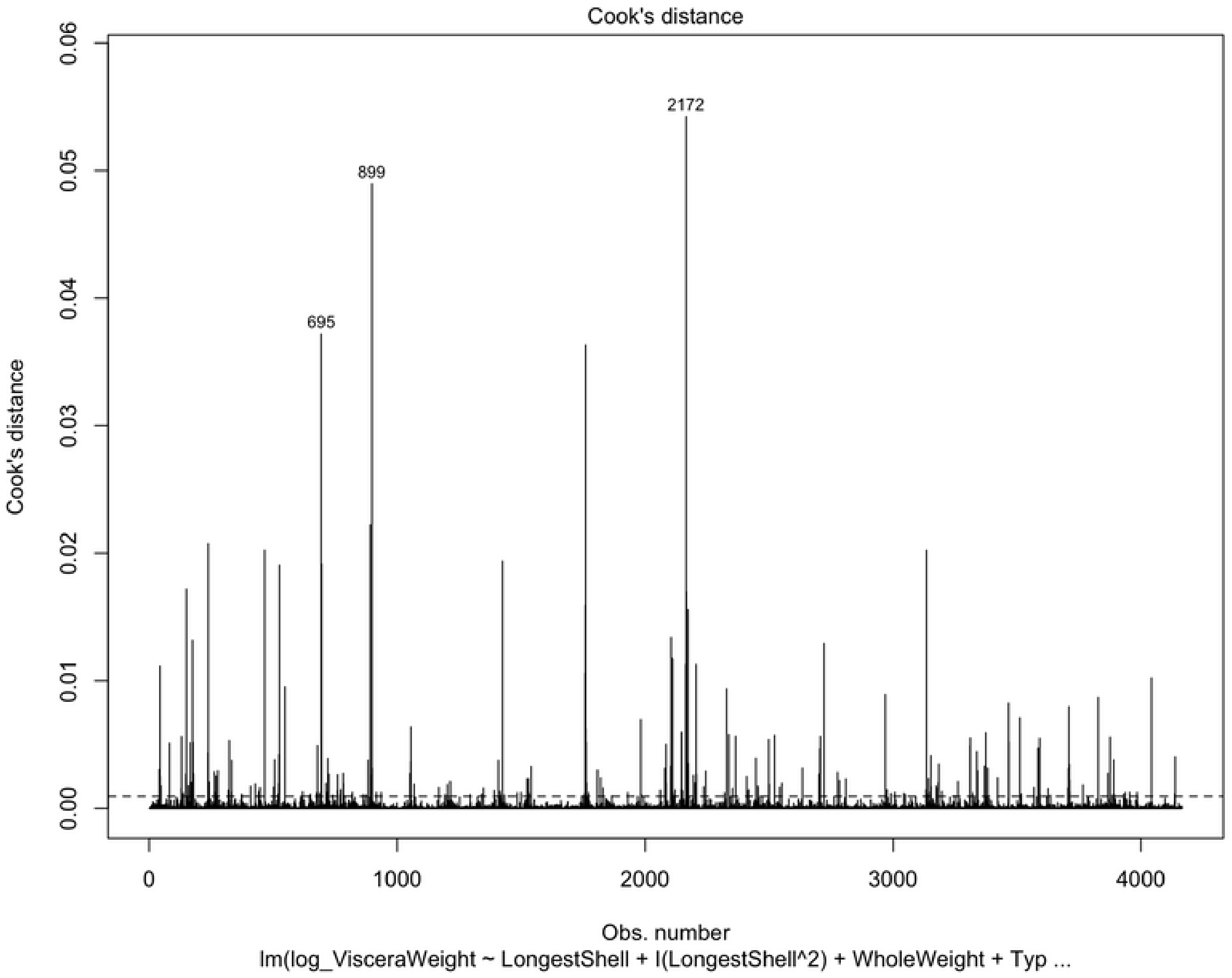


Multiple Linear Regression

log_tran

- **Viscera Weight** : Residual Diagnostic

plot(abalone.step,which=4)



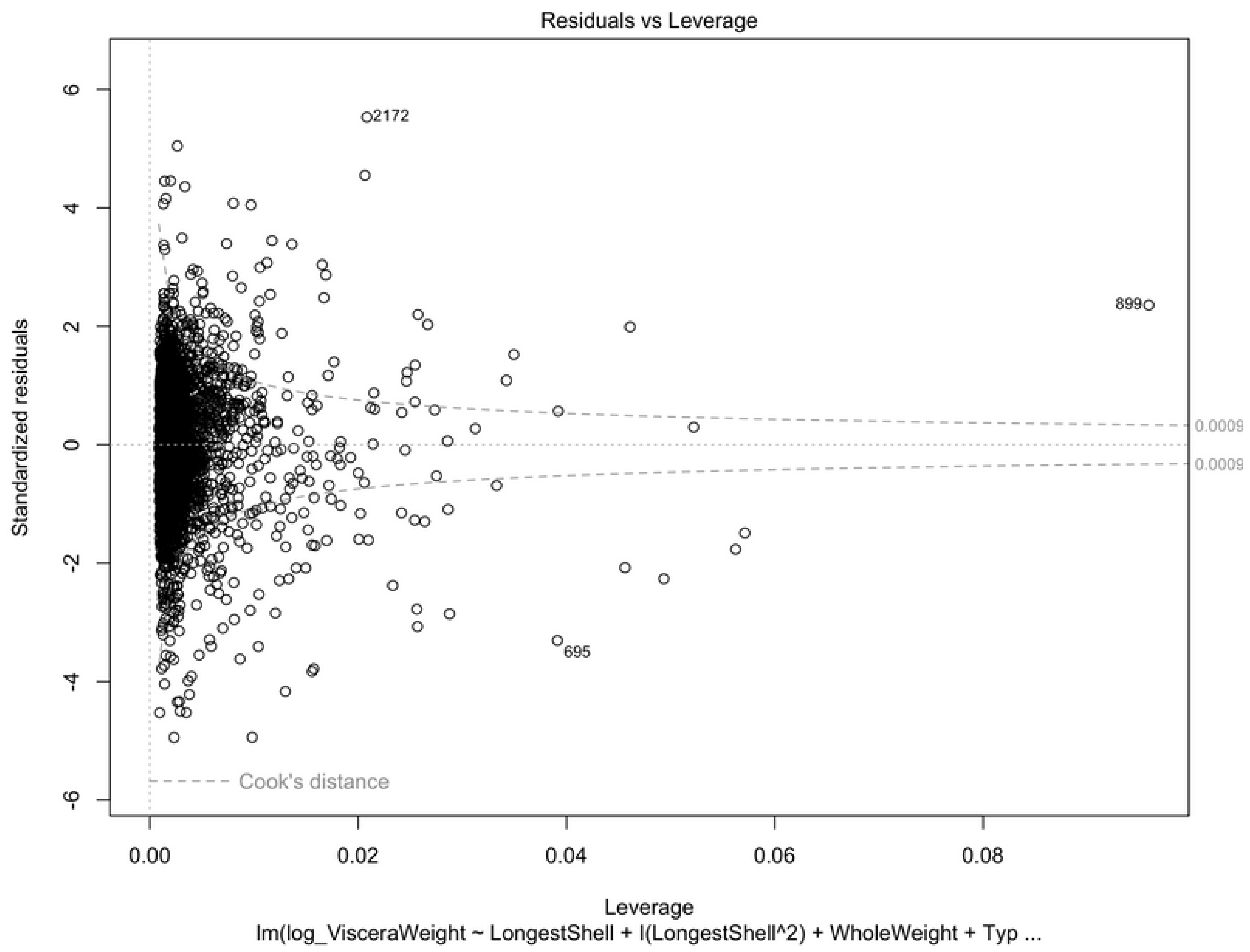
Multiple Linear Regression

log_tran

04

- **Viscera Weight** : Residual Diagnostic

plot(abalone.step,which=5)



Multiple Linear Regression

log_tran

- Viscera Weight

General equation for fitted model is :

ln(VisceraWeight) =

$$\boxed{\begin{aligned}&-7.499 + 9.002(LongestShell) - 6.966(LongestShell^2)\\&+ 1.302(WholeWeight) - 0.1751(WholeWeight^2)\\&+ 26.65(Height) - 160.7(Height^2) + 316.9(Height^3)\\&+ 3.042(Diameter) - 4.062(Diameter^2)\\&- 0.05554(TypeI) -0.01641(TypeM)\end{aligned}}$$

from, ln(x)=y ⟶ x=e^y

VisceraWeight = e ^

Conclusion
ANOVA

Shucked Weight

จากผลการวิเคราะห์ข้อมูล จะได้ค่า P-value < 2.2e-16 < 0.05 จึงสรุปได้ว่า Reject H0 = Accept H1 ที่ระดับความเชื่อมั่น 95% ดังนั้นค่าเฉลี่ยน้ำหนักเนื้อของหอยเป๋าฮื้อ (ShuckedWeight) ในแต่ละ sexTypeนั้น แตกต่างกัน

Viscera Weight

จากผลการวิเคราะห์ข้อมูล จะได้ค่า P-value < 2.2e-16 < 0.05 จึงสรุปได้ว่า Reject H0 = Accept H1 ที่ระดับความเชื่อมั่น 95% ดังนั้นค่าเฉลี่ยน้ำหนักอวัยวะเพศของหอยเป๋าฮื้อ (VisceraWeight) ในแต่ละTypeนั้น แตกต่างกัน

Conclusion
Interval Estimate

Shucked Weight

จากผลการวิเคราะห์ Interval Estimate ที่ความเชื่อมั่น 90% และ 95% พบว่า เพศเมีย(Type F) มีน้ำหนักเฉลี่ยในส่วนของเนื้อหอยเป๋าฮื้อมากที่สุด

Viscera Weight

จากผลการวิเคราะห์ Interval Estimate ที่ความเชื่อมั่น 95% พบว่า เพศเมีย(Type F) มีน้ำหนักเฉลี่ยในส่วนของอวัยวะภายในหอยเป๋าฮื้อมากที่สุด

Conclusion MLR
Multiple Linear Regression

Shucked Weight

จากผลการวิเคราะห์ข้อมูล จะได้ค่า P-value < 2.2e-16 < 0.05 จึงสรุปได้ว่า Reject H0 = Accept H1 ที่ระดับความเชื่อมั่น 95% ดังนั้นตัวแปรอิสระทุกตัวสามารถร่วมกันทำนายพยากรณ์ตัวแปรตามได้ จึงได้สมการทำนายShuckedWeight ออกมา

Viscera Weight

จากผลการวิเคราะห์ข้อมูล จะได้ค่า P-value < 2.2e-16 < 0.05 จึงสรุปได้ว่า Reject H0 = Accept H1 ที่ระดับความเชื่อมั่น 95% ดังนั้นตัวแปรอิสระทุกตัวสามารถร่วมกันทำนายพยากรณ์ตัวแปรตามได้ จึงได้สมการทำนายVisceraWeight ออกมา

How about ?

แบ่งตามช่วงอายุ

Dataset details

06

Abalone

Age = rings + 1.5

sex	Length (mm)	Diameter (mm)	Height (mm)	Whole. weight (g)	Shucked. weight (g)	Viscera. weight (g)	Shell. weight (g)	rings
F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

Age.cat 06

- Child
- Adult
- Old

```
> abalone <- within(abalone, {
+   Age.cat <- NA
+   Age.cat[Age < 6] <- "Child"
+   Age.cat[Age >= 6 & Age < 14] <- "Adult"
+   Age.cat[Age >= 14] <- "Old"
+ } ) ;head(abalone)
  Type LongestShell Diameter Height WholeWeight ShuckedWeight VisceraWeight
1    M        0.455    0.365  0.095      0.5140        0.2245        0.1010
2    M        0.350    0.265  0.090      0.2255        0.0995        0.0485
3    F        0.530    0.420  0.135      0.6770        0.2565        0.1415
4    M        0.440    0.365  0.125      0.5160        0.2155        0.1140
5    I        0.330    0.255  0.080      0.2050        0.0895        0.0395
6    I        0.425    0.300  0.095      0.3515        0.1410        0.0775
  ShellWeight Rings  Age Age.cat
1       0.150    15 16.5     Old
2       0.070     7  8.5   Adult
3       0.210     9 10.5   Adult
4       0.155    10 11.5   Adult
5       0.055     7  8.5   Adult
6       0.120     8  9.5   Adult
```

Shucked Weight

06

```
> var(abalone[abalone$Age.cat == 'Child',c('VisceraWeight')])
[1] 6.881303e-05
> var(abalone[abalone$Age.cat == 'Adult',c('VisceraWeight')])
[1] 0.01145955
> var(abalone[abalone$Age.cat == 'Old',c('VisceraWeight')])
[1] 0.009309416
> # We will see that these variance are not equal to each other.
> one_way2 <- oneway.test(ShuckedWeight~Age.cat, data = abalone, var.equal = FALSE)
> one_way2 # the mean of wholeweight are not equal.

	One-way analysis of means (not assuming equal variances)

data:  ShuckedWeight and Age.cat
F = 4090.5, num df = 2.0, denom df = 1169.6, p-value < 2.2e-16
```

Viscera Weight 06

```
> var(abalone[abalone$Age.cat == 'Child',c('ShuckedWeight')])
[1] 0.0002646575
> var(abalone[abalone$Age.cat == 'Adult',c('ShuckedWeight')])
[1] 0.04924526
> var(abalone[abalone$Age.cat == 'Old',c('ShuckedWeight')])
[1] 0.0368474
> # We will use Welch ANOVA.
> one_way1 <- oneway.test(VisceraWeight~Age.cat, data = abalone, var.equal = FALSE)
> one_way1 # the mean of wholeweight are not equal.

	One-way analysis of means (not assuming equal variances)

data:  VisceraWeight and Age.cat
F = 4329.9, num df = 2.0, denom df = 1115.6, p-value < 2.2e-16
```

Interval Estimate 06

- ## Shucked Weight

(Two Means t–test two–sided)

ดังนั้น Old มีน้ำหนักส่วนเนื้อหอยที่มากที่สุด
ที่ระดับความเชื่อมั่นที่ 95%

Age.cat	Hypothesis	P-value	Confidence interval
Adult, Child	H0 : μA - μC = 0 H1 : μA - μC ≠ 0	p-value < 2.2e-16	95% CI: (0.3220115, 0.3386653)
Old, Child	H0 : μO - μC = 0 H1 : μO - μC ≠ 0	p-value < 2.2e-16	95% CI: (0.3962053, 0.4257818)
Old, Adult	H0 : μO - μA = 0 H1 : μO - μA ≠ 0	p-value < 2.2e-16	95% CI: (0.06452124, 0.09678894)

Interval Estimate 06

- **Viscera Weight**

(Two Means t-test two-sided)

ดังนั้น Old มีน้ำหนักส่วนอวัยวะภายในที่มากที่สุด ที่ระดับความเชื่อมั่นที่ 95%

Age.cat	Hypothesis	P-value	Confidence interval
Adult, Child	H0 : μA - μC = 0 H1 : μA - μC ≠ 0	p-value < 2.2e-16	95% CI: (0.1571349, 0.1652625)
Old, Child	H0 : μO - μC = 0 H1 : μO - μC ≠ 0	p-value < 2.2e-16	95% CI: (0.2196980, 0.2345779)
Old, Adult	H0 : μO - μA = 0 H1 : μO - μA ≠ 0	p-value < 2.2e-16	95% CI: (0.05789808, 0.07398040)

Multiple Linear Regression

- ## ตัวแปรสุ่มที่ต้องการศึกษา

Shucked Weight

ตัวแปรต้น : ความยาวของเปลือกหอย(Length), เส้นผ่านศูนย์กลางของเปลือกหอย(Diameter),ความสูงของเปลือกหอย(Height),น้ำหนักทั้งหมดของหอยเป๋าฮื้อ(Whole.weight),ช่วงอายุ(Age)
ตัวแปรตาม : น้ำหนักเนื้อหอยเป๋าฮื้อ (Shucked.weight)

Viscera Weight

ตัวแปรต้น : ความยาวของเปลือกหอย(Length), เส้นผ่านศูนย์กลางของเปลือกหอย(Diameter),ความสูงของเปลือกหอย(Height),น้ำหนักทั้งหมดของหอยเป๋าฮื้อ(Whole.weight),ช่วงอายุ(Age)
ตัวแปรตาม : น้ำหนักอวัยวะภายในเนื้อหอยเป๋าฮื้อ (Viscera.weight)

Multiple Linear Regression

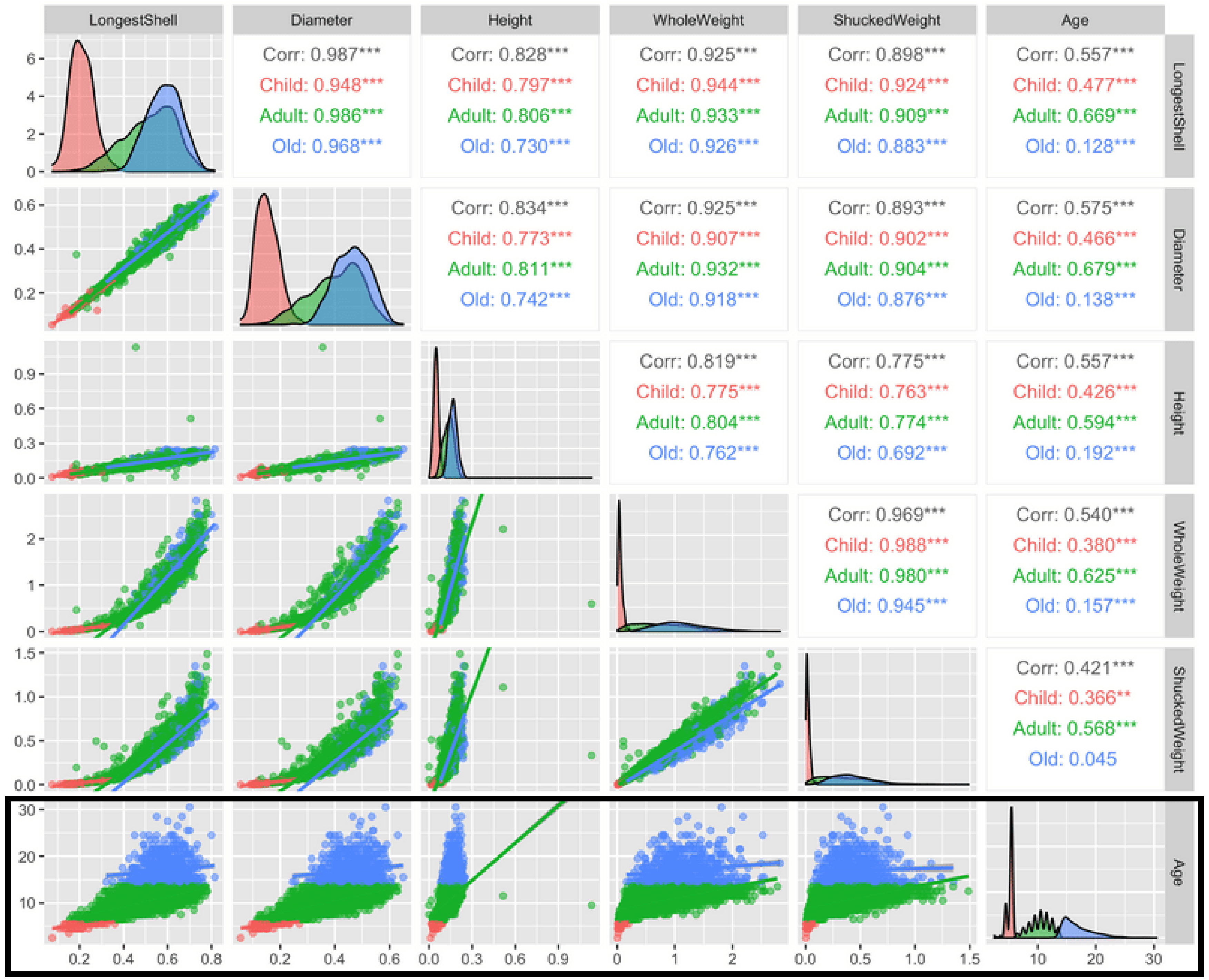
- **สมมติฐาน**

H0 : ตัวแปรอิสระทุกตัวไม่สามารถร่วมกันทำนายพยากรณ์ตัวแปรตามได้ (No effect)

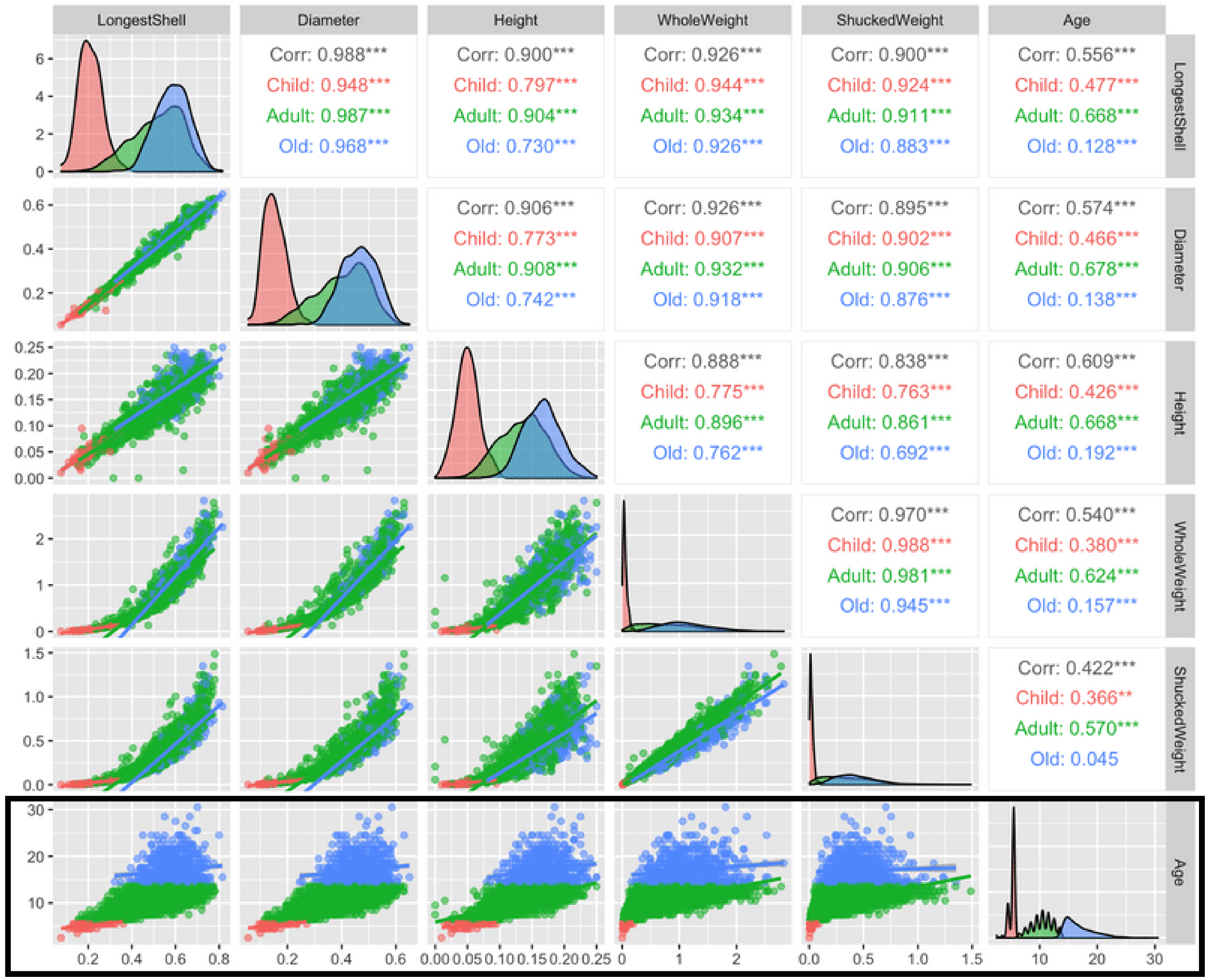
H1 : ตัวแปรอิสระทุกตัวสามารถร่วมกันทำนายพยากรณ์ตัวแปรตามได้ (At the least one ≠ 0)

Multiple Linear Regression

- Age



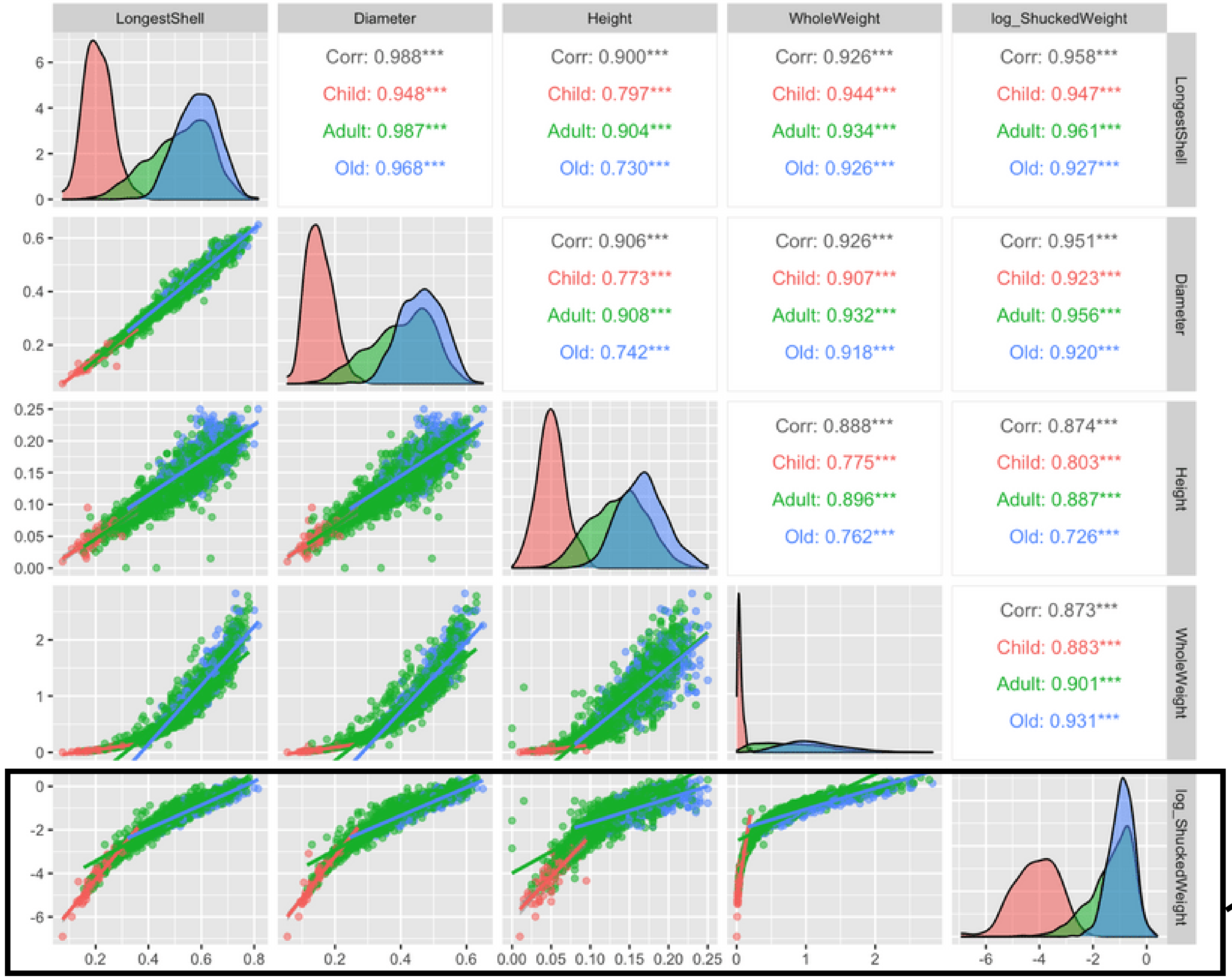
Plot with outlier



Plot without outlier

Multiple Linear Regression

- **Shucked Weight**



Log Transformation

Multiple Linear Regression

- **Shucked Weight**

summary >>

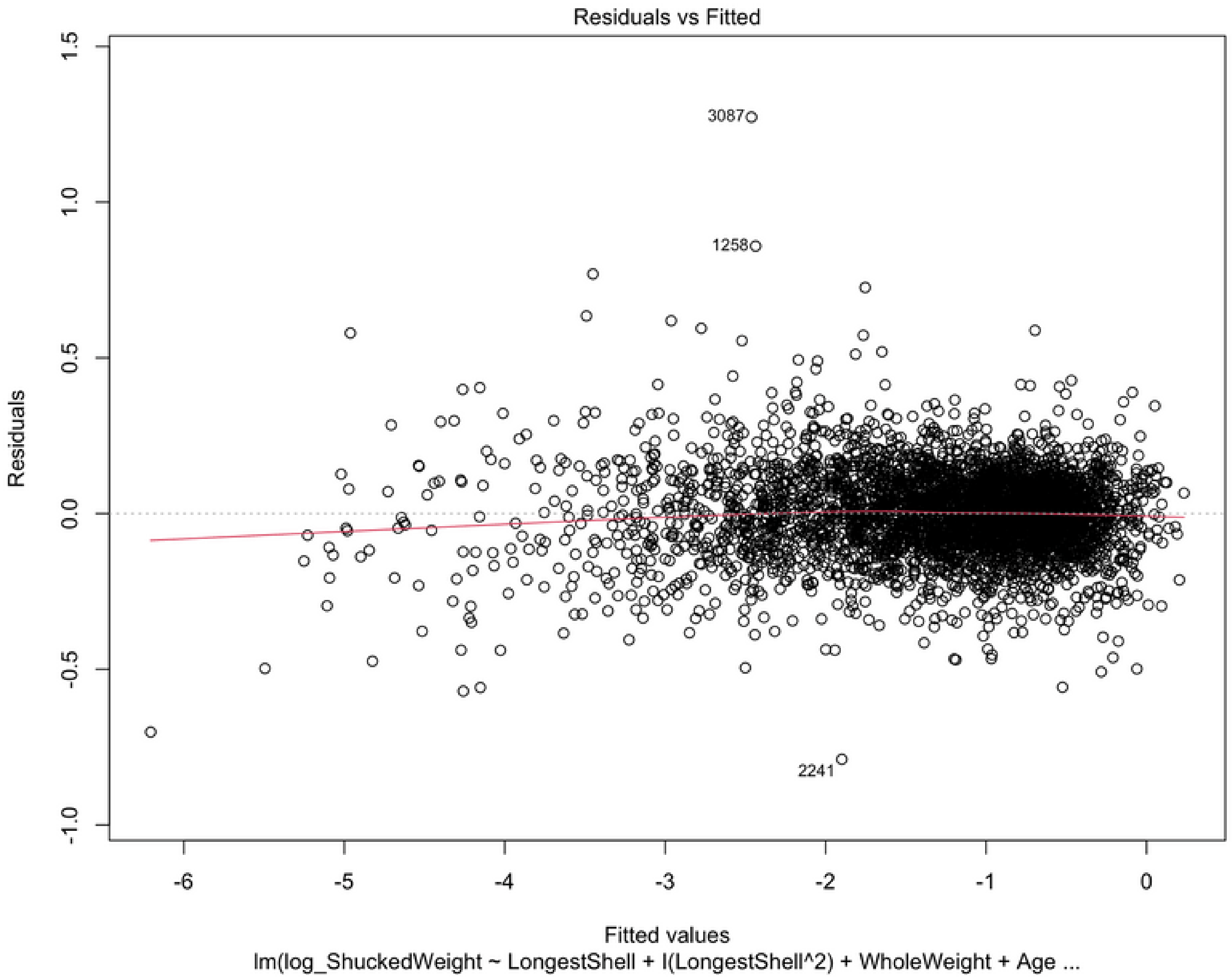
```
Call:
lm(formula = log_ShuckedWeight ~ LongestShell + I(LongestShell^2) +
    WholeWeight + Age.cat + I(WholeWeight^2) + I(Height^2) +
    Height + I(Height^3) + I(Diameter^2) + Diameter + WholeWeight:Age.cat,
    data = abalone0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.78929 -0.07896  0.00206  0.07864  1.27302

Coefficients:
                           Estimate Std. Error  t value Pr(>|t|)
(Intercept)                -7.35826    0.04100 -179.450  < 2e-16 ***
LongestShell               12.03333    0.61349   19.614  < 2e-16 ***
I(LongestShell^2)         -10.03922    0.54336  -18.476  < 2e-16 ***
WholeWeight                 6.00625    0.45875   13.093  < 2e-16 ***
Age.catAdult                0.42004    0.03481   12.066  < 2e-16 ***
Age.catOld                  0.29966    0.03822    7.840 5.69e-15 ***
I(WholeWeight^2)           -0.21428    0.01405  -15.251  < 2e-16 ***
I(Height^2)              -128.50630    9.58557  -13.406  < 2e-16 ***
Height                     16.94744    1.27902   13.250  < 2e-16 ***
I(Height^3)               281.72355   23.22741   12.129  < 2e-16 ***
I(Diameter^2)              -3.43955    0.80281   -4.284 1.87e-05 ***
Diameter                    2.68371    0.70999    3.780 0.000159 ***
WholeWeight:Age.catAdult   -4.35200    0.45456   -9.574  < 2e-16 ***
WholeWeight:Age.catOld     -4.37566    0.45457   -9.626  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1354 on 4158 degrees of freedom
Multiple R-squared:  0.9754,	Adjusted R-squared:  0.9753
F-statistic: 1.266e+04 on 13 and 4158 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

• Shucked Weight

Residual Diagnostic >>

plot(abalone.step,which=1)

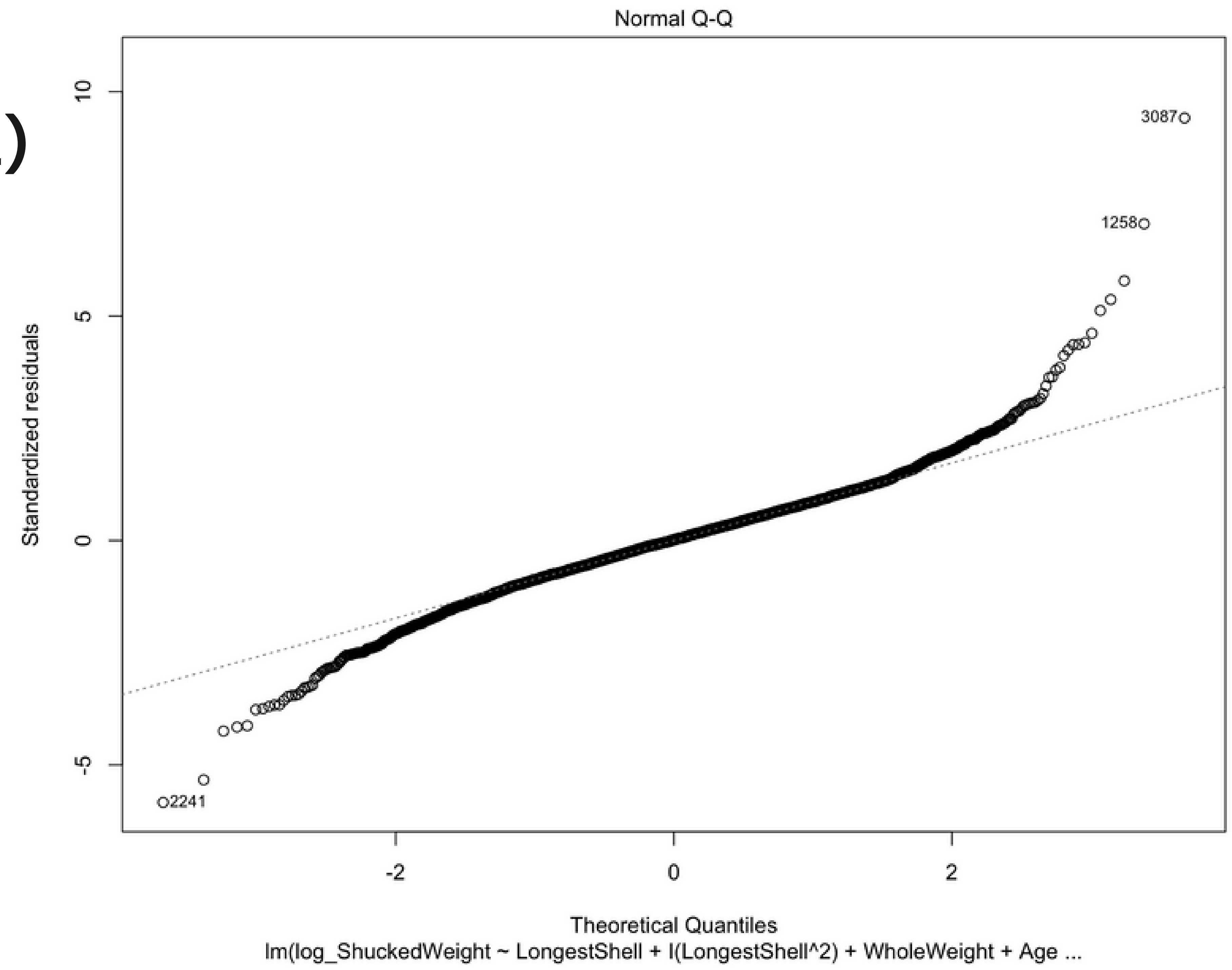


Multiple Linear Regression

- **Shucked Weight**

Residual Diagnostic >>

plot(abalone.step,which=2)

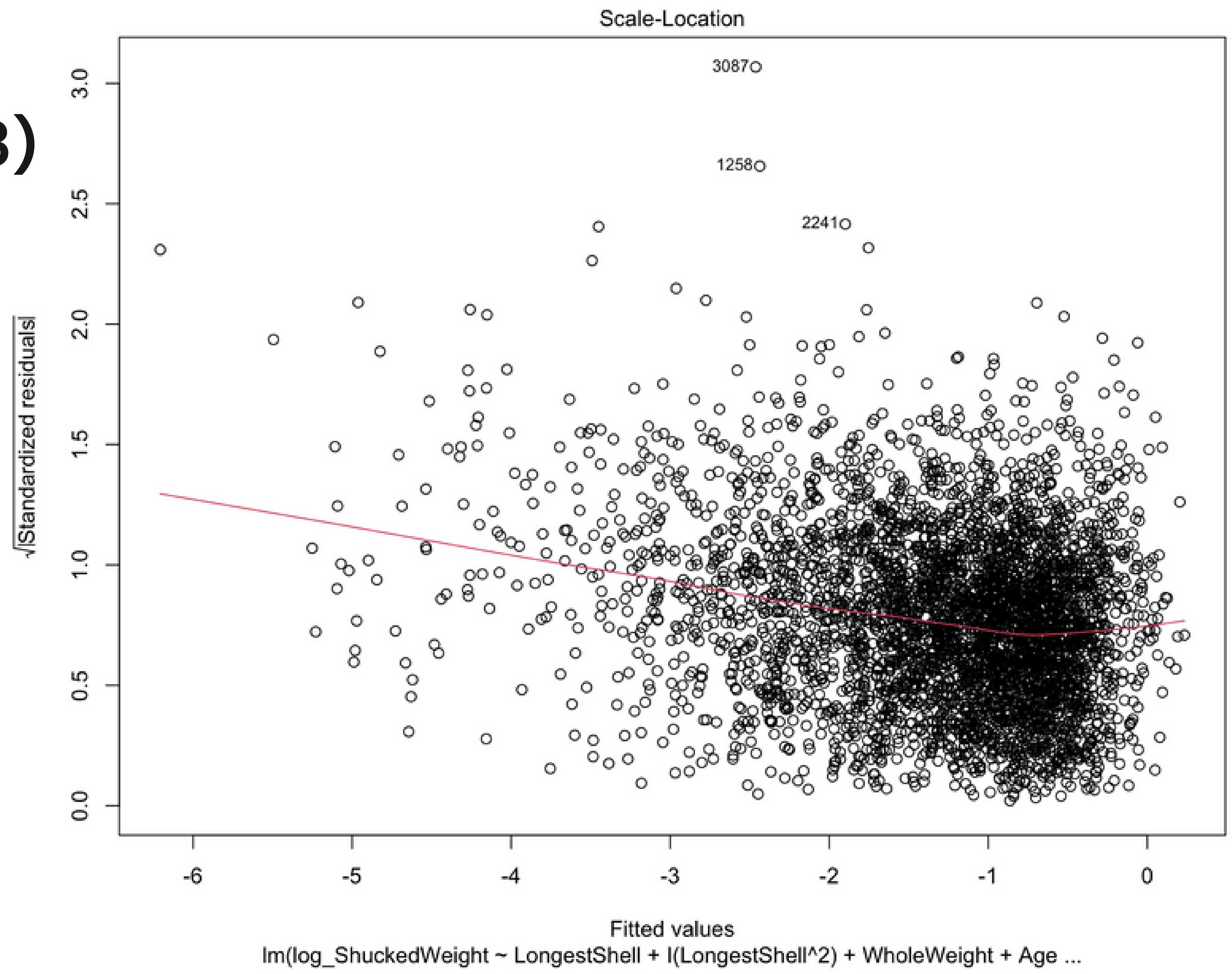


Multiple Linear Regression

- **Shucked Weight**

Residual Diagnostic >>

plot(abalone.step,which=3)

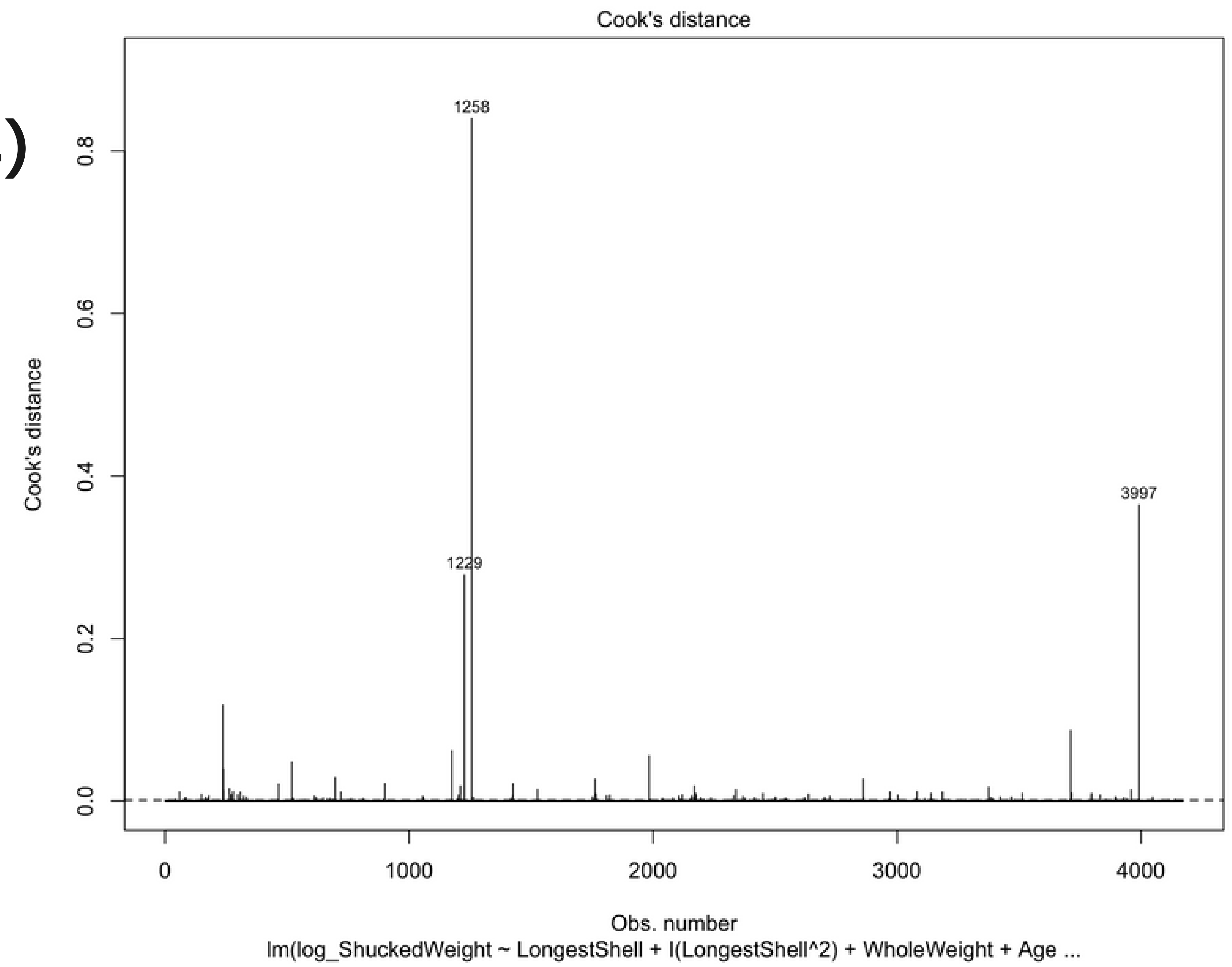


Multiple Linear Regression

- **Shucked Weight**

Residual Diagnostic >>

plot(abalone.step,which=4)

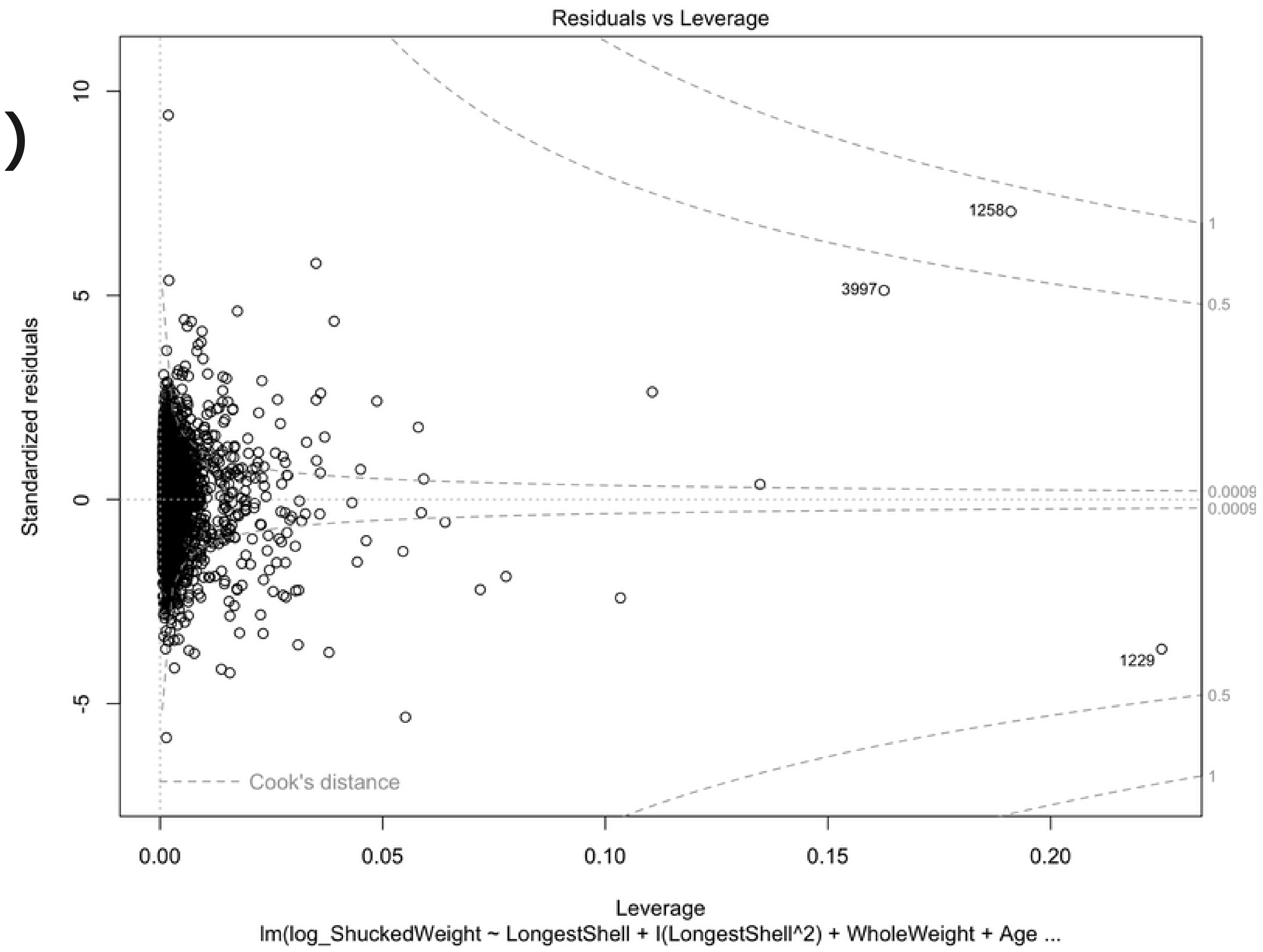


Multiple Linear Regression

- **Shucked Weight**

Residual Diagnostic >>

plot(abalone.step,which=5)



Multiple Linear Regression

• Shucked Weight

General equation for fitted model is :

ln(ShuckedWeight) =

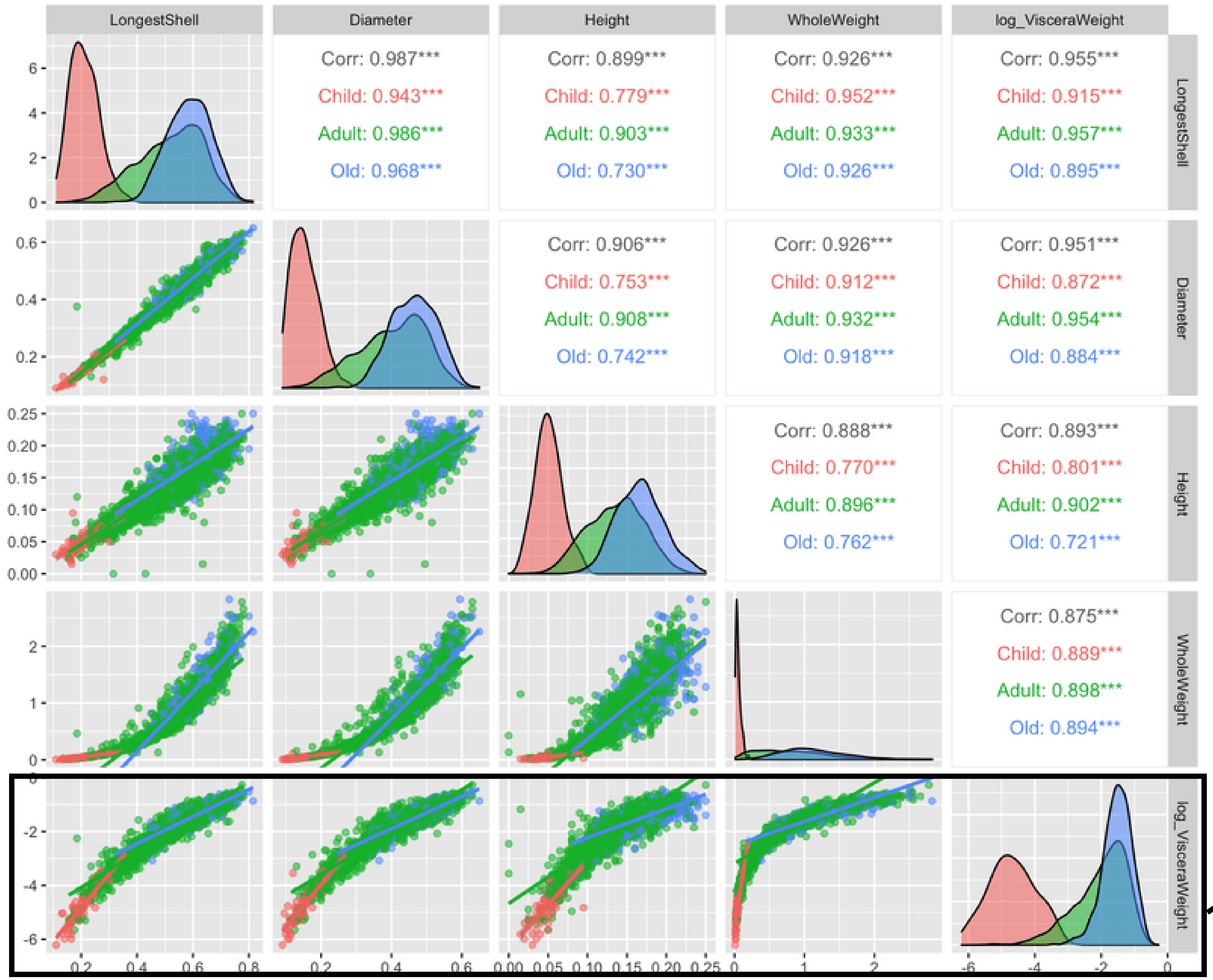
-7.35826 + 12.03333(LongestShell) - 10.03922(LongestShell^2)
+ 6.00625(WholeWeight) - 0.21428(WholeWeight^2)
+ 16.94744(Height) - 128.50630(Height^2) + 281.72355(Height^3)
+ 2.68371(Diameter) - 3.43955(Diameter^2)
+ 0.42004(Age.catAdult) + 0.29966(Age.catOld)
-4.35200(WholeWeight:Age.catAdult)
-4.37566(WholeWeight:Age.catOld)

from, log(x)=y ⟶ x=10^y

ShuckedWeight = e ^

Multiple Linear Regression

- **Viscera Weight**



Log Transformation

Multiple Linear Regression

- **Viscera Weight**

summary >>

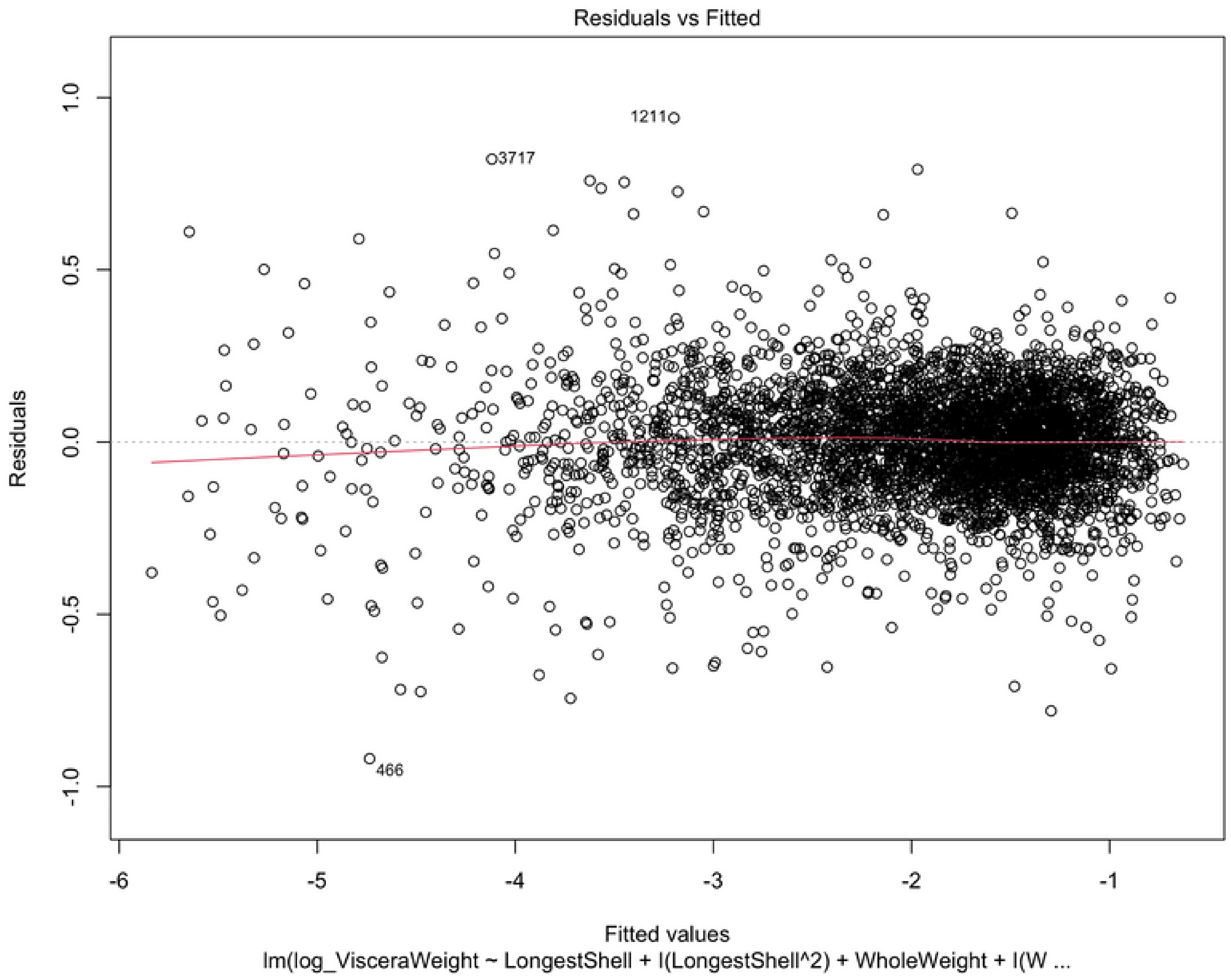
```
Call:
lm(formula = log_VisceraWeight ~ LongestShell + I(LongestShell^2) +
    WholeWeight + I(WholeWeight^2) + Age.cat + Height + I(Height^2) +
    I(Height^3) + I(Diameter^2) + Diameter + WholeWeight:Age.cat,
    data = abalone0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.91933 -0.09149  0.00377  0.09534  0.94098

Coefficients:
                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)               -7.37165    0.04964 -148.497  < 2e-16 ***
LongestShell               6.67765    0.63741   10.476  < 2e-16 ***
I(LongestShell^2)         -5.25010    0.57773   -9.087  < 2e-16 ***
WholeWeight                5.37007    0.54791    9.801  < 2e-16 ***
I(WholeWeight^2)          -0.20057    0.01667  -12.029  < 2e-16 ***
Age.catAdult               0.45474    0.04144   10.972  < 2e-16 ***
Age.catOld                 0.47254    0.04557   10.369  < 2e-16 ***
Height                    12.67601    1.52407    8.317  < 2e-16 ***
I(Height^2)              -68.82848   11.43040   -6.022 1.88e-09 ***
I(Height^3)              124.86357   27.70429    4.507 6.76e-06 ***
I(Diameter^2)             -7.29726    0.86605   -8.426  < 2e-16 ***
Diameter                   6.21724    0.75306    8.256  < 2e-16 ***
WholeWeight:Age.catAdult  -3.96229    0.54314   -7.295 3.56e-13 ***
WholeWeight:Age.catOld    -4.00136    0.54322   -7.366 2.11e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1611 on 4158 degrees of freedom
Multiple R-squared:  0.9634,    Adjusted R-squared:  0.9633
F-statistic:  8417 on 13 and 4158 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

- **Viscera Weight**

Residual Diagnostic >>

plot(abalone.step,which=1)

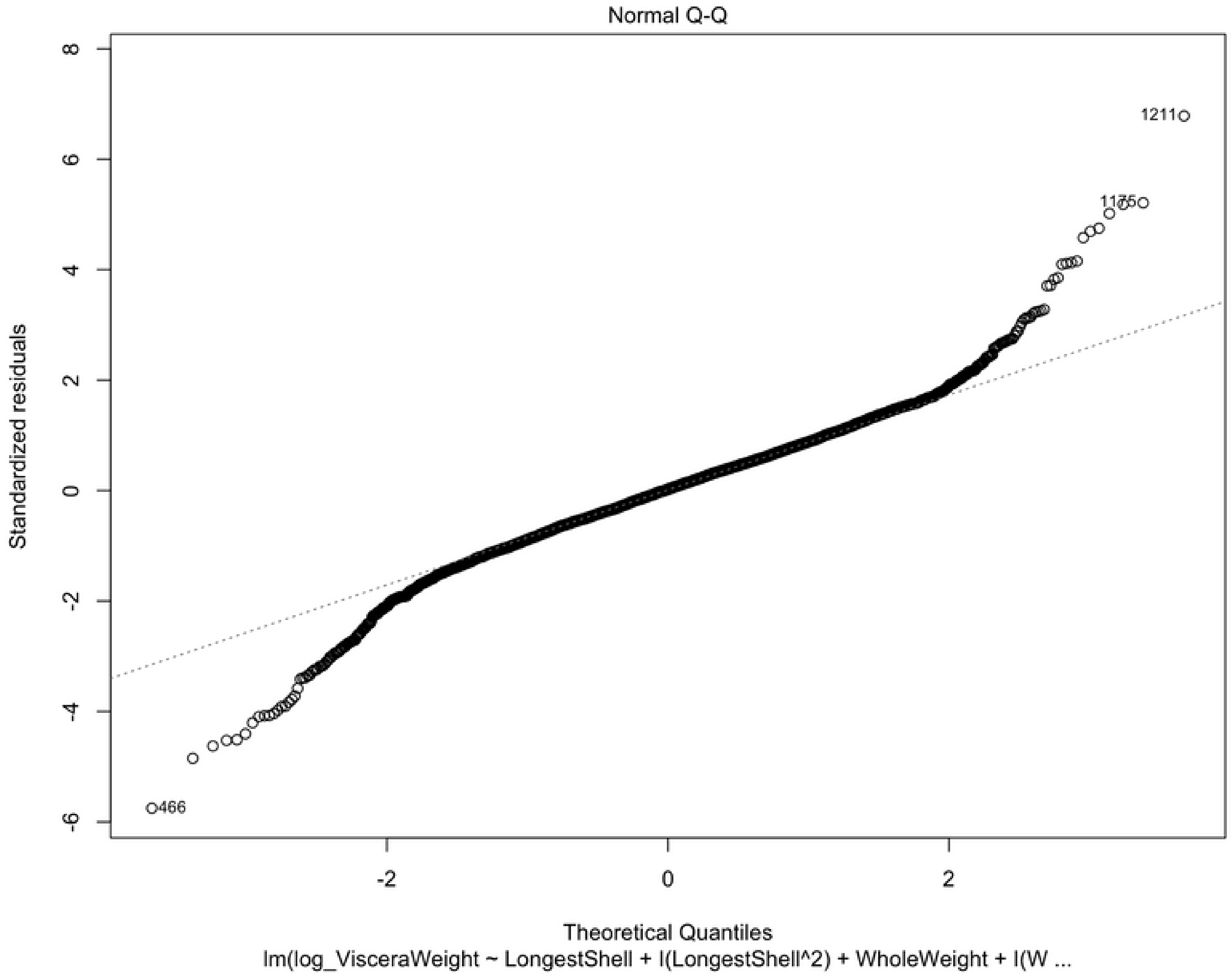


Multiple Linear Regression

• Viscera Weight

Residual Diagnostic >>

plot(abalone.step,which=2)

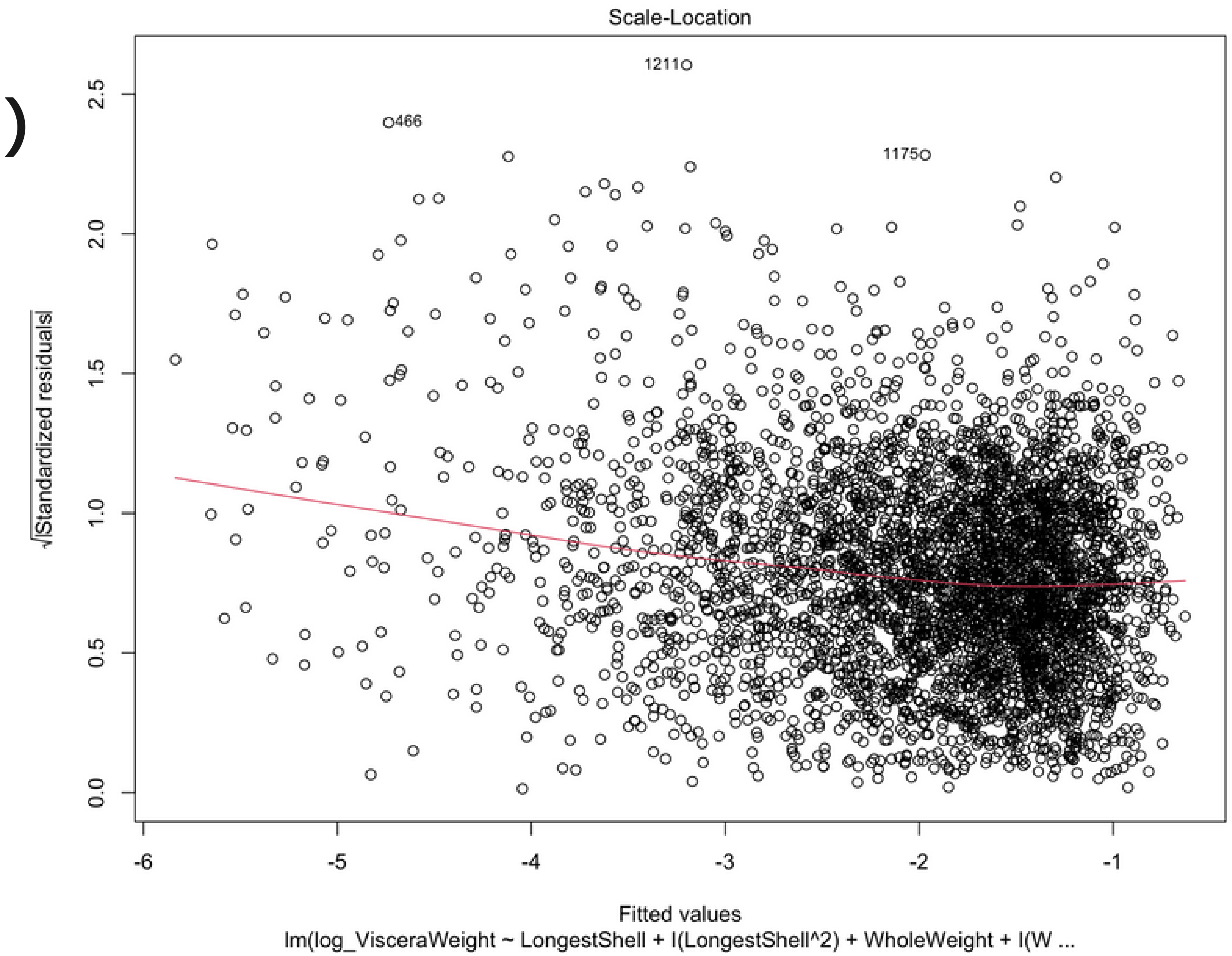


Multiple Linear Regression

- **Viscera Weight**

Residual Diagnostic >>

plot(abalone.step,which=3)

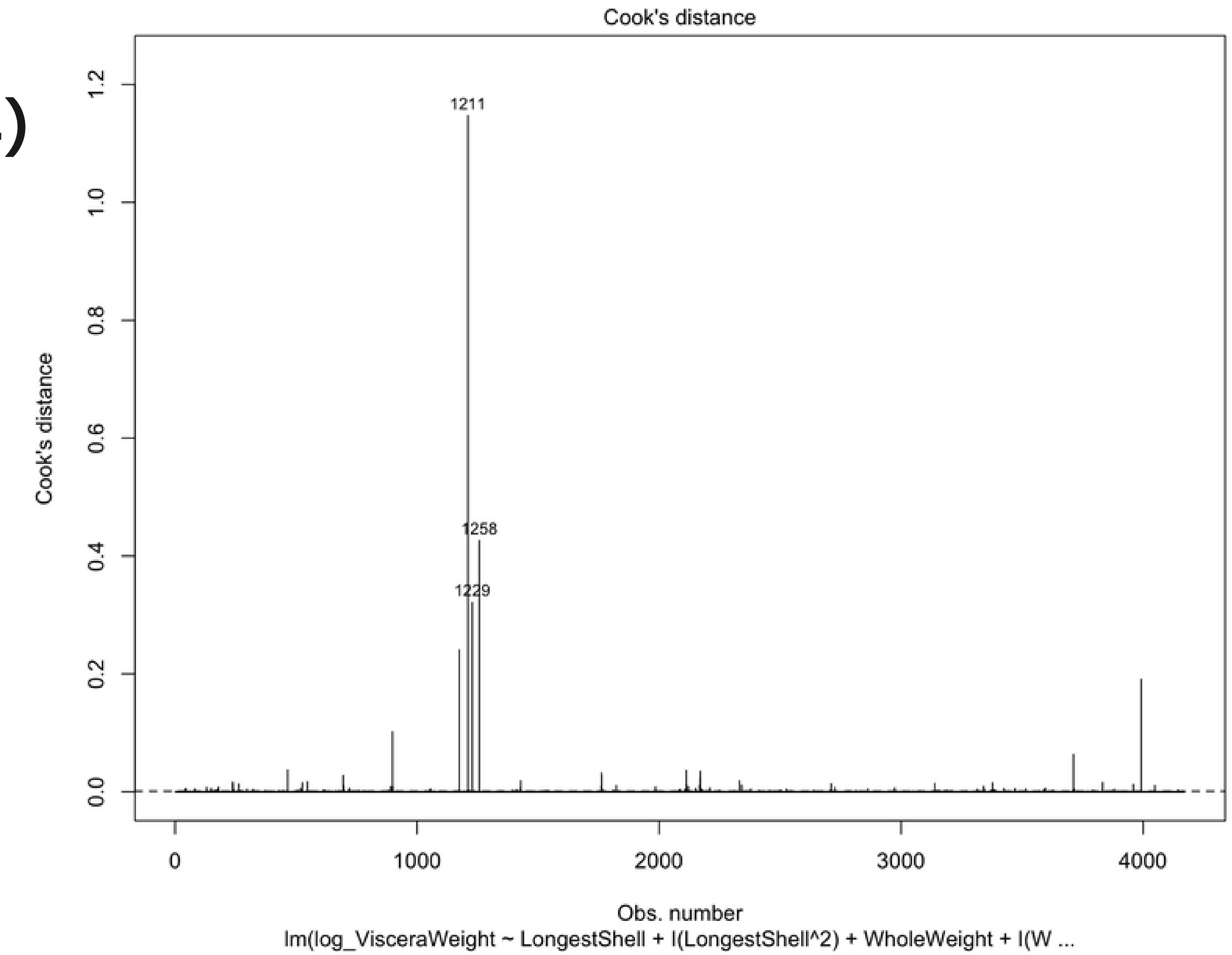


Multiple Linear Regression

- **Viscera Weight**

Residual Diagnostic >>

plot(abalone.step,which=4)

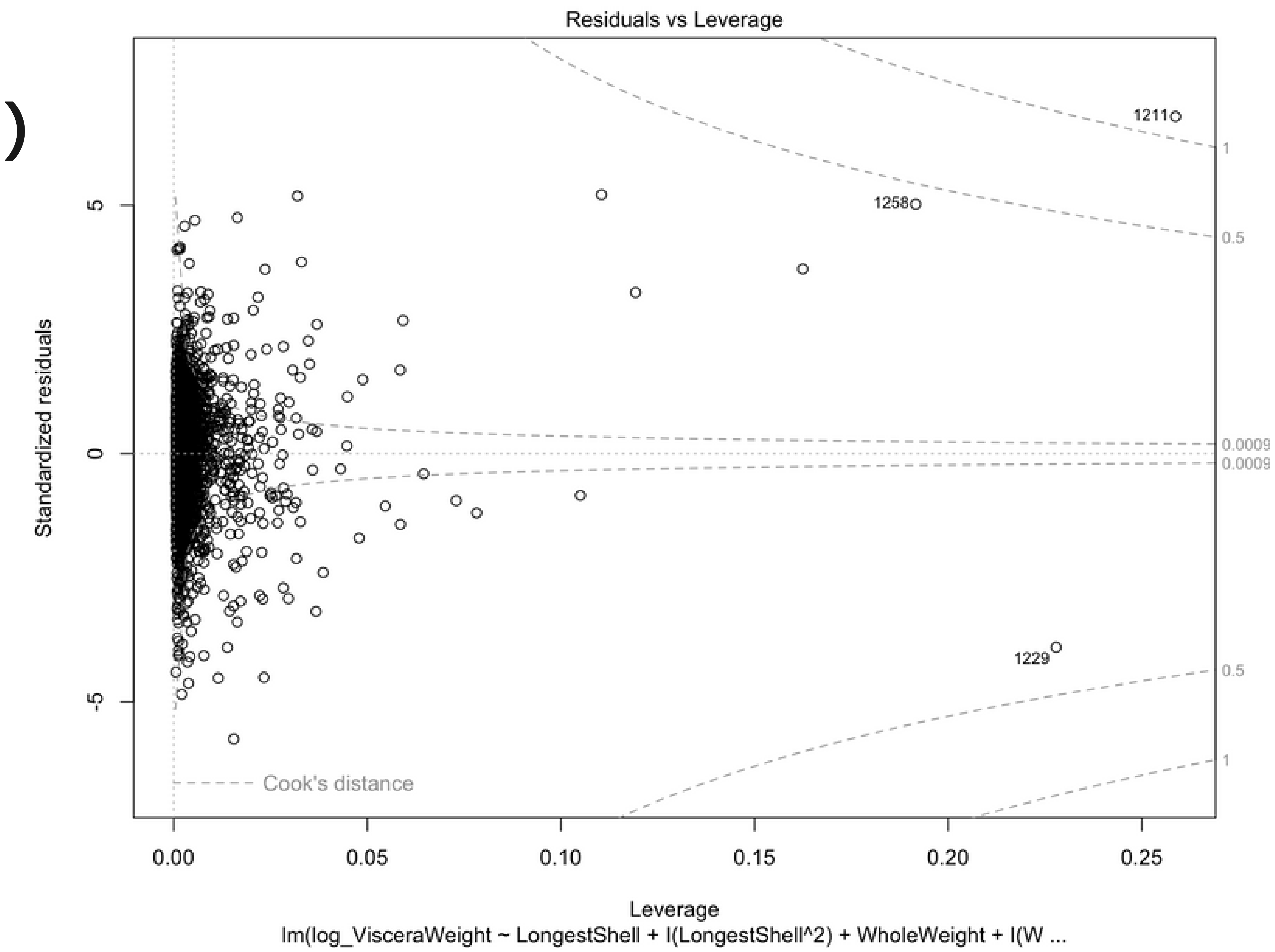


Multiple Linear Regression

- **Viscera Weight**

Residual Diagnostic >>

plot(abalone.step,which=5)



Multiple Linear Regression

- **Viscera Weight**

General equation for fitted model is :

$$
\ln(\text{VisceraWeight}) = -7.37165 + 6.67765(\text{LongestShell}) - 5.25010(\text{LongestShell}^2) + 5.37007(\text{WholeWeight}) - 0.20057(\text{WholeWeight}^2) + 12.67601(\text{Height}) - 68.82848(\text{Height}^2) + 124.86357(\text{Height}^3) + 6.21724(\text{Diameter}) - 7.29726(\text{Diameter}^2) + 0.45474(\text{Age.catAdult}) + 0.47254(\text{Age.catOld}) - 3.96229(\text{WholeWeight:Age.catAdult}) - 4.00136(\text{WholeWeight:Age.catOld})
$$

from, $\log(x)=y \longrightarrow x=10^y$

VisceraWeight = e ^

Reference

Project Abalone

01 Chapter8-Sampling Distribution and Confidence Intervals (Slide)

02 Chapter9-Hypothesis Testing (Slide)

03 Chapter10-Analysis of variance : ANOVA (Slide)

04 Chapter12-Multiple Linear Regression (Slide)

05 Chapter13-Linear Model Selection and Diagonostics (Slide)

06 Inhibition of SARS-CoV-2 Virus Entry by the Crude Polysaccharides of Seaweeds and Abalone Viscera In Vitro (Research Paper)