

Residual Analysis

S. Eanes

11/25/2020

Setup data

```
coral <- read.csv("C:/Users/shado/Documents/School/Fossil Coral/data/coral_3weighted.csv")

#omit data with no response
coral <- coral[!is.na(coral$U238),]

#select three largest species
genus_trim <- c("Acropora", "Porites")
#remove coral with age > 10
coral <- coral[which(coral$Genus %in% genus_trim),] %>% dplyr::filter(Age < 10)
coral <- coral[(coral$Calcite <= 1 | is.na(coral$Calcite)),]

#clean up a nice dataframe
coral.df <- coral %>% mutate(Temperature = Temp) %>%
  select(U238,pH,TAlk,Salinity,Temperature,OmegaA,TCO2) %>%
  data.frame()

coral.df$U238 <- coral.df$U238 * 0.421
nrow(coral.df)
```

```
## [1] 700
```

```
#This is a favia, so removing favia removes it
#coral[92,]
#coral[coral$Source==74,]
##obs 92 from site 74 is an extreme outlier
#coral.df <- coral.df[-92,]
```

```
## the next two outliers are 773 and 787 in coral.df, or 774 and 788 in coral
coral[c(114,647,661),]
```

| ## | Lon | Lat | ID | Source | Calcite | U238 | U238sig | Age |
|--------|-----------|-----------|----------|--------------------|--------------|--------------------|------------|--------------|
| ## 146 | -157.4527 | 1.94412 | 2531 | 89 | 0.5 | 0.254141 | 0.00037562 | 3.916663 |
| ## 684 | 151.9491 | -23.44794 | 3575 | 118 | NA | 6.564800 | 0.0039 | 4.477378 |
| ## 698 | 151.9491 | -23.44794 | 3595 | 118 | NA | 5.910700 | 0.0071 | 4.927027 |
| ## | U234delta | Basin | Genus | | | Place | | Location |
| ## 146 | 146.2019 | Pacific | Porites | | Line Islands | Kiritimati | | Line Islands |
| ## 684 | 144.6395 | Pacific | Acropora | Great Barrier Reef | Heron Reef | Great Barrier Reef | | |

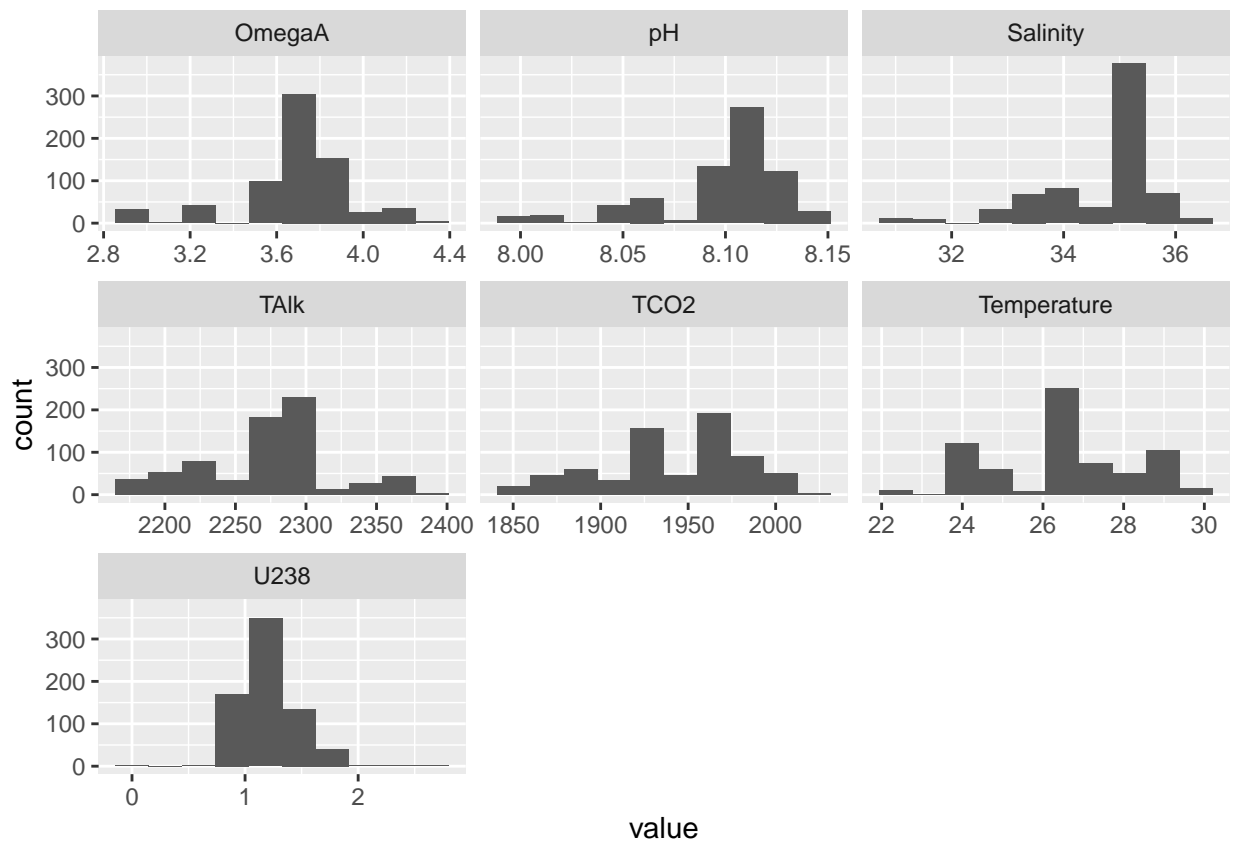
```
## 698 144.1144 Pacific Acropora Great Barrier Reef, Heron Reef Great Barrier Reef
##           Site LabID      Temp Salinity      pH      TAlk      TC02      OmegaA
## 146 Kiritimati      3 26.85735 35.08406 8.059197 2292.901 1985.586 3.530822
## 684 Heron Reef      25 24.39374 35.46241 8.099540 2301.872 1966.979 3.698729
## 698 Heron Reef      25 24.39374 35.46241 8.099540 2301.872 1966.979 3.698729
##           oxygen
## 146 201.3794
## 684 219.1517
## 698 219.1517
```

#collected at different sites, from the same lab

Data Distribution

Plot the distributions of the predictors.

```
ggplot(gather(coral.df), aes(value)) +
  geom_histogram(bins=10) +
  facet_wrap(~key, scales = 'free_x')
```



Residual Analysis

Residuals vs fitted looks perfectly normal - no signs of non-linearity or heteroskedasticity.

Normal Q-Q is somewhat concerning. According the shapiro test we reject the null hypothesis and conclude the data is non-normal.

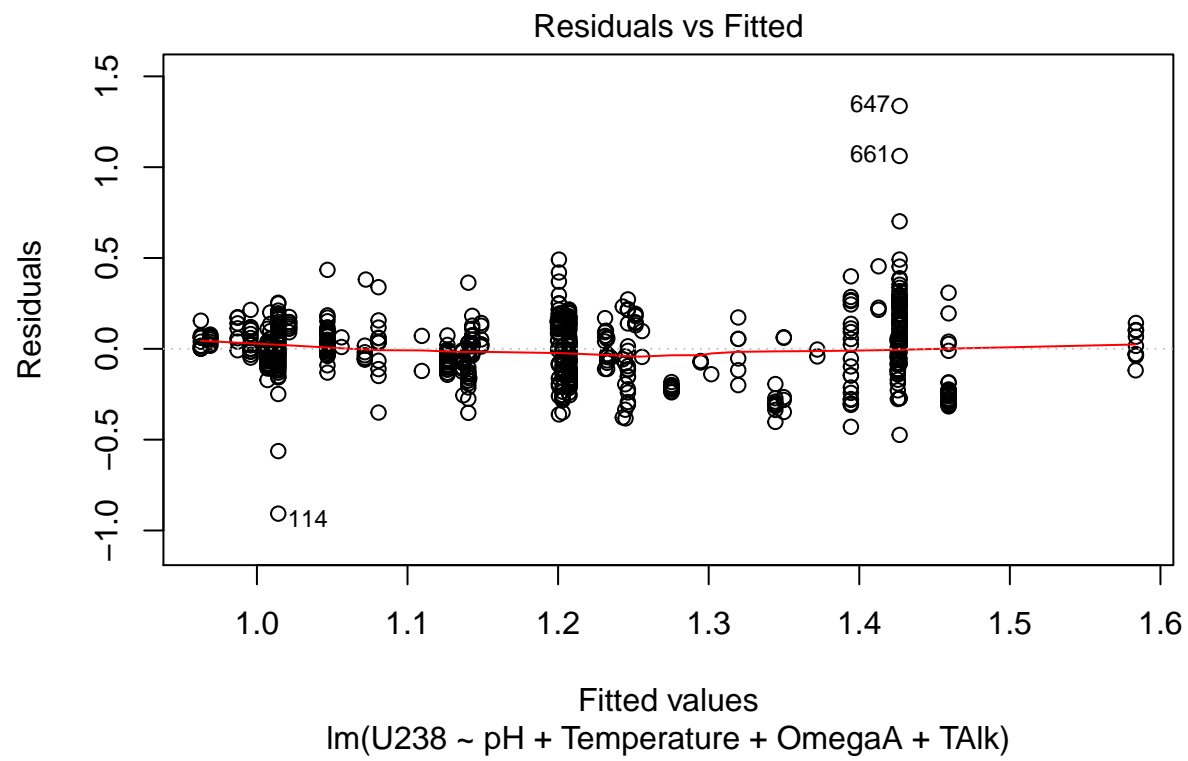
Scale-location is maybe slightly non-linear, but not to an extent that is cause for concern. We can also formally test heteroskedasticity with a Breusch-Pagan test. The null hypothesis is homoskedasticity, and we fail to reject it.

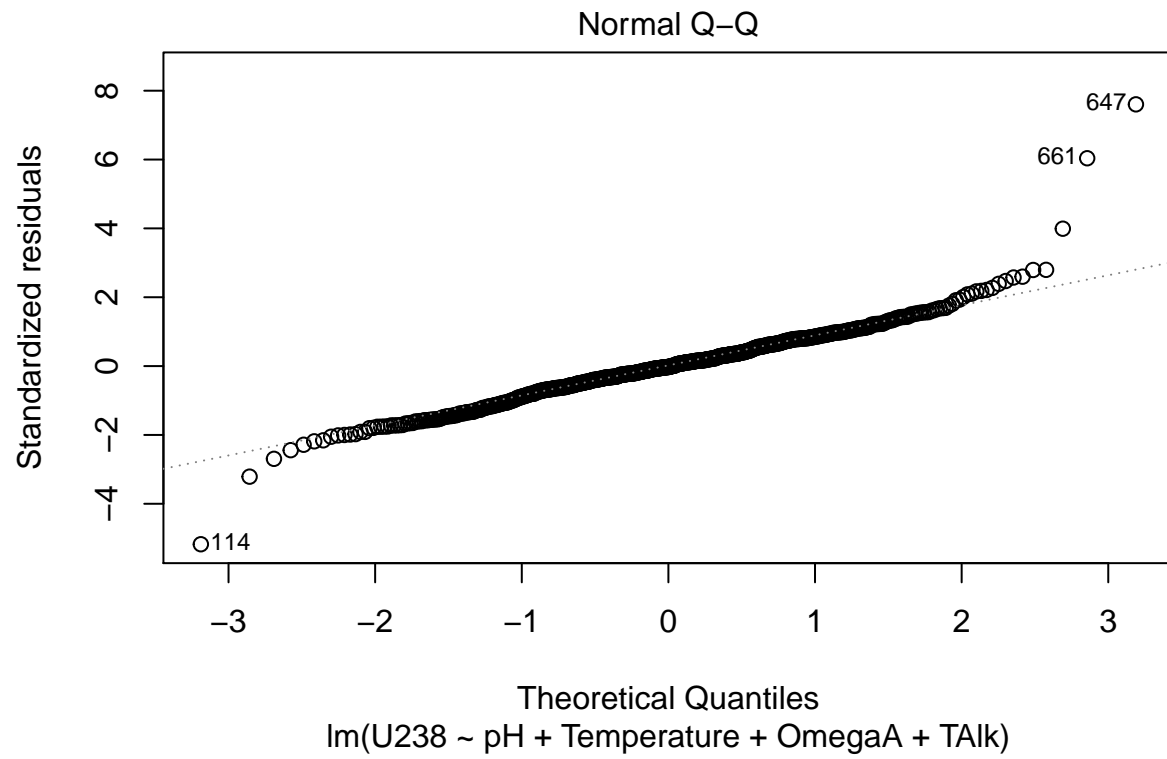
There are no concerning observations when looking at cook's distance in our residuals vs leverage plot.

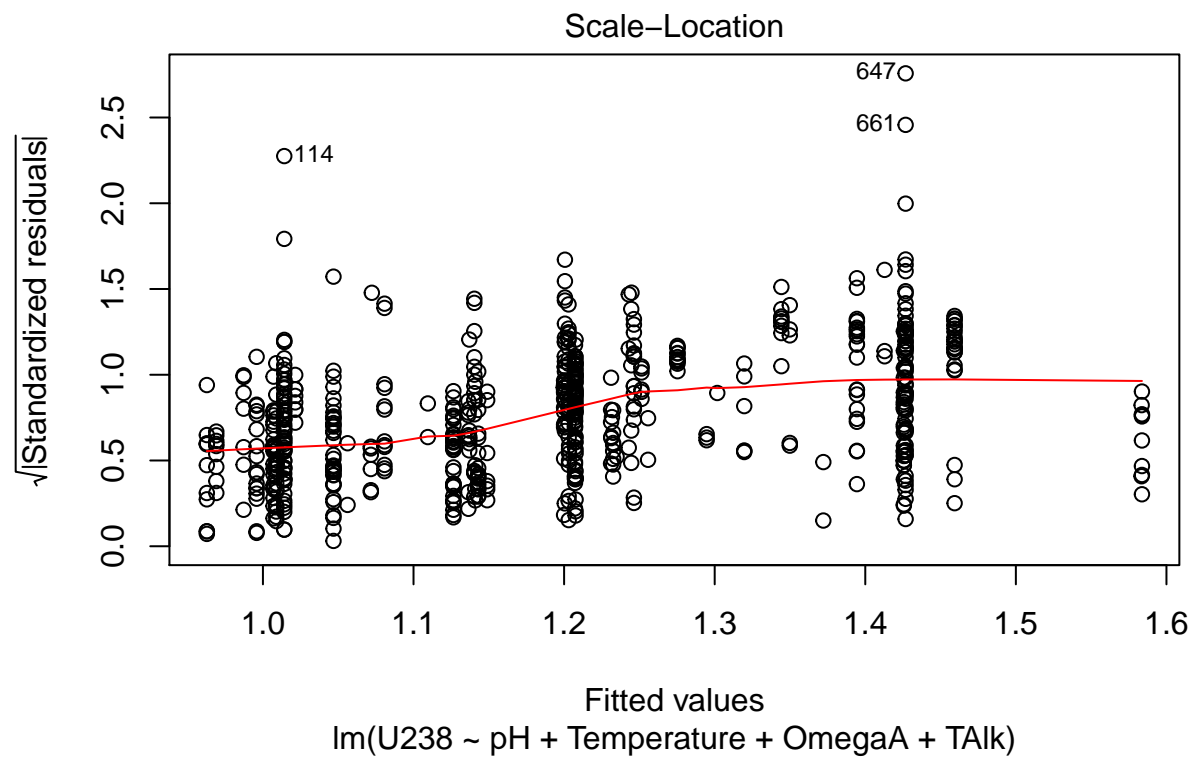
```
options(scipen=999)
final_model <- lm(U238 ~ pH + Temperature + OmegaA + TAlk, data=coral.df)
summary(final_model)
```

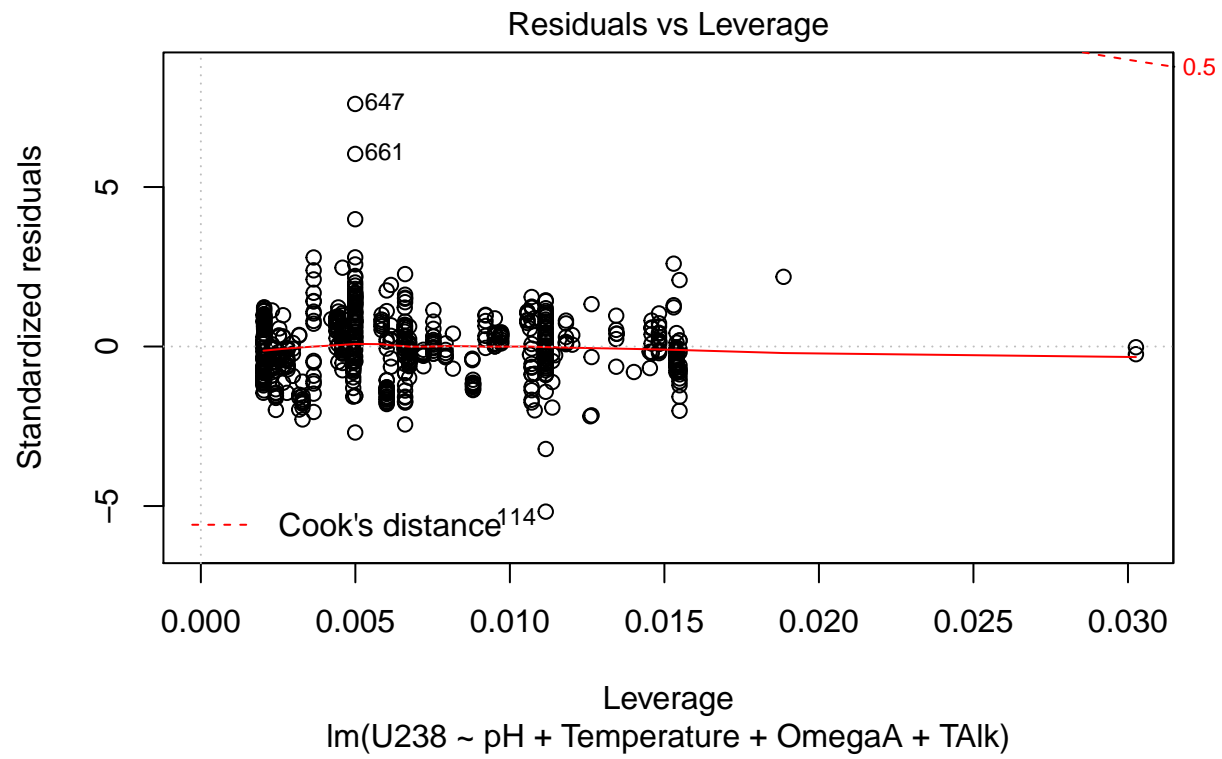
```
##
## Call:
## lm(formula = U238 ~ pH + Temperature + OmegaA + TAlk, data = coral.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90719 -0.09916 -0.00437  0.10731  1.33697
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -84.5117096  12.9166557  -6.543 0.000000000011712163 ***
## pH           10.4282870   1.5292630   6.819 0.00000000001991412 ***
## Temperature  -0.0620973   0.0075820  -8.190 0.000000000000000125 ***
## OmegaA       -1.1252224   0.2199593  -5.116 0.00000040502659236 ***
## TAlk          0.0031065   0.0005718   5.433 0.00000007673312738 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1763 on 695 degrees of freedom
## Multiple R-squared:  0.4383, Adjusted R-squared:  0.4351
## F-statistic: 135.6 on 4 and 695 DF,  p-value: < 0.00000000000000022
```

```
plot(final_model)
```



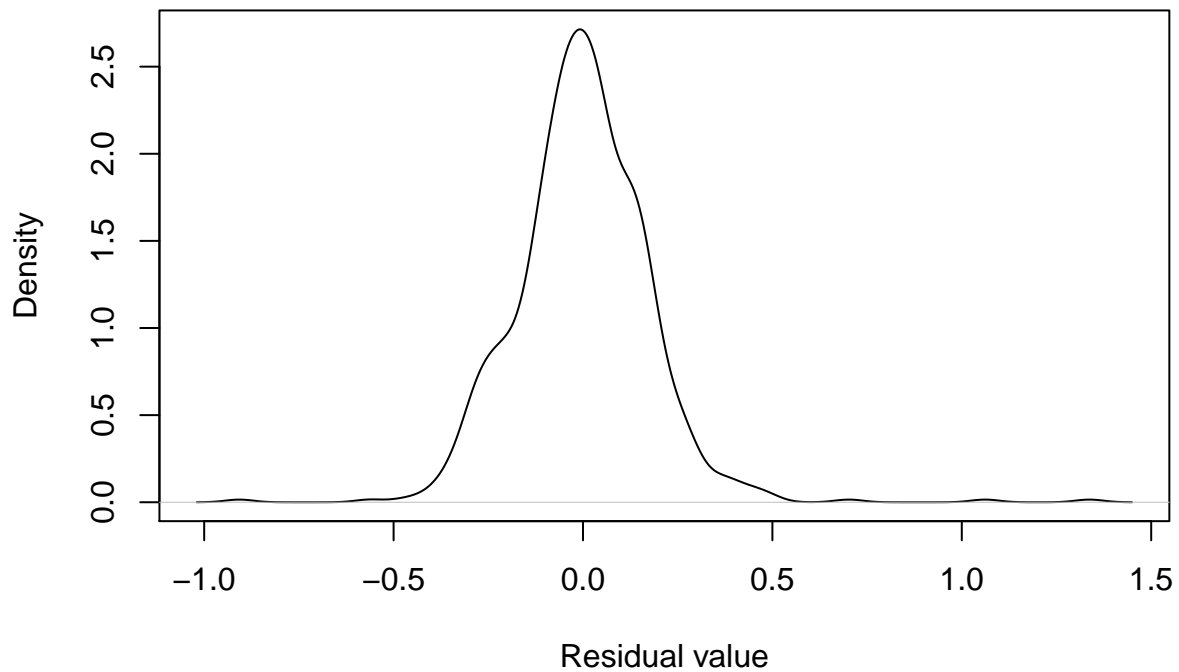






```
#normal Q-Q additional analysis
par(mfrow=c(1,1))
d<-density(final_model[['residuals']])
plot(d,main='Residual KDE Plot',xlab='Residual value')
```

Residual KDE Plot



```
shapiro.test(final_model[['residuals']])
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  final_model[["residuals"]]  
## W = 0.94198, p-value = 0.0000000000000006668
```

```
#scale-location additional analysis  
bptest(final_model)
```

```
##  
##  studentized Breusch-Pagan test  
##  
## data:  final_model  
## BP = 27.896, df = 4, p-value = 0.00001309
```