

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, U238=U238*.421) %>%
  select(U238,pH,TAlk,Salinity,Temperature,OmegaA,TCO2) %>%
  data.frame()

nrow(coral.df)
```

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

```
coral.summary =coral %>% mutate(Temperature = Temp, U238=U238*.421) %>%
  select(U238,Age,Calcite,U234delta,pH,TAlk,Salinity,Temperature,OmegaA,TCO2) %>%
  data.frame()

summary <- coral.summary %>%
  gather %>% group_by(key) %>%
  summarise("5%"=quantile(value,.05,na.rm=T),
            median=median(value,na.rm=T),
            mean=mean(value,na.rm=T),
            "95%"=quantile(value,.95,na.rm=T)) %>%
  as.data.frame()
cbind(c(summary[,1]),round(summary[,-1],2))
```

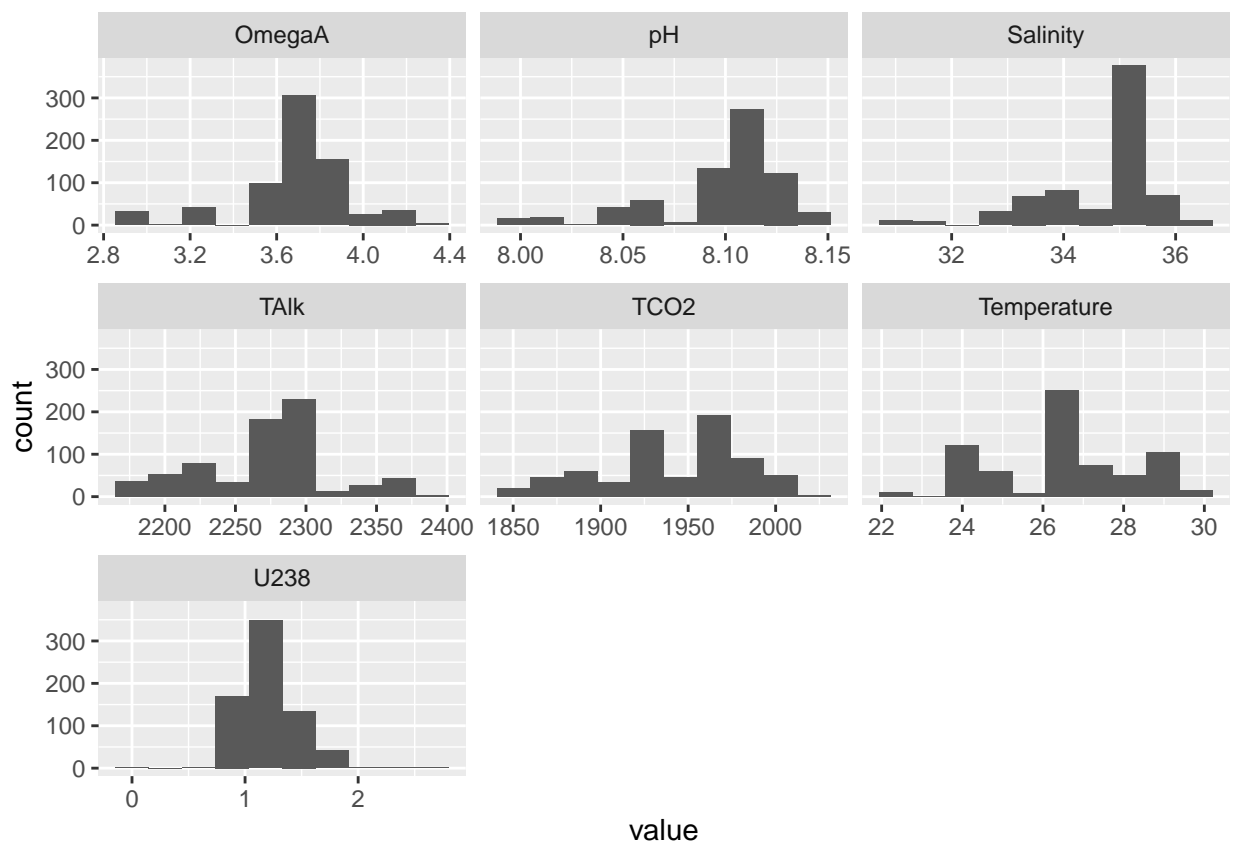
##	c(summary[, 1])	5%	median	mean	95%
## 1	Age	0.01	0.46	2.40	9.00
## 2	Calcite	0.00	0.50	0.44	1.00
## 3	OmegaA	3.16	3.76	3.69	4.18

## 4	pH	8.05	8.11	8.10	8.13
## 5	Salinity	33.05	35.11	34.72	35.94
## 6	TAIk	2180.89	2264.88	2272.45	2364.90
## 7	TCO2	1870.91	1953.53	1943.54	2001.25
## 8	Temperature	24.29	26.65	26.63	29.35
## 9	U234delta	141.23	144.83	144.66	148.51
## 10	U238	0.93	1.14	1.20	1.66

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.

As can be seen in the normal q-q plot, there is some non-normality. In the KDE plot below, and with the shapiro test we can confirm that there is some right skew. However, with a large number of samples like this analysis, linear regression is quite robust to violations of the normality assumption. Furthermore, by excluding high percent calcite samples, we are confident that the samples in this analysis are of high quality.

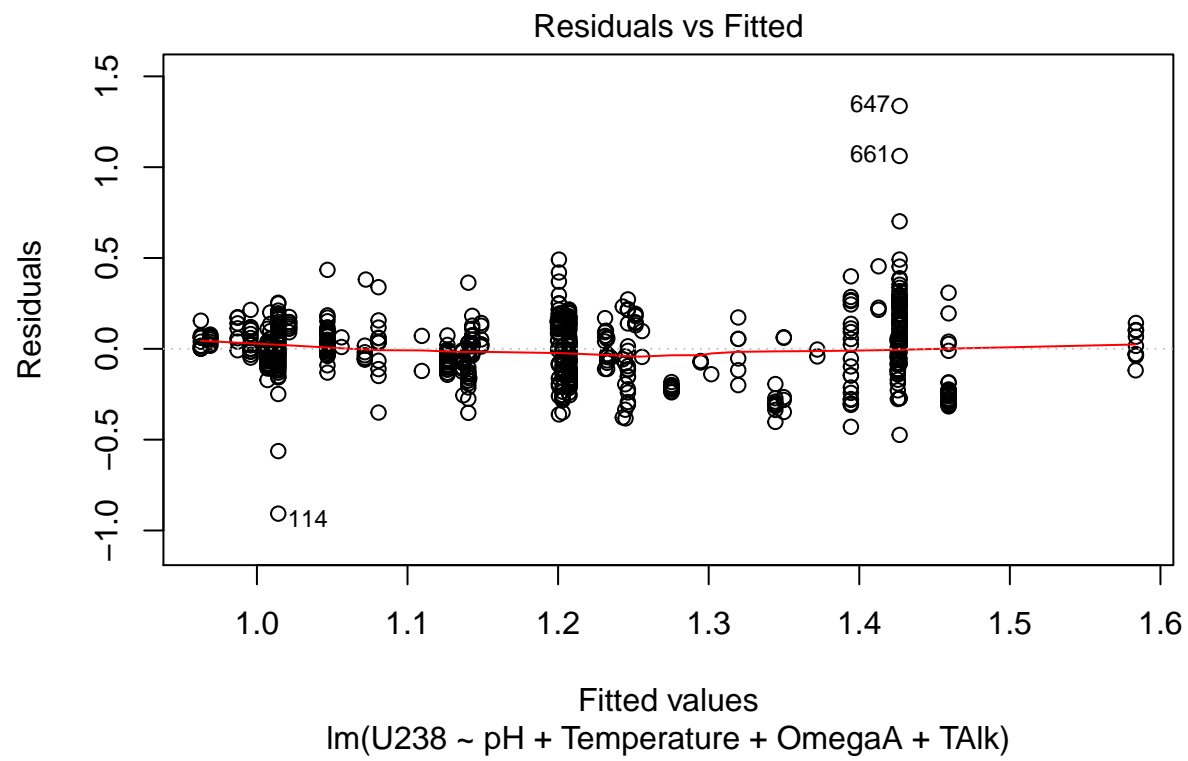
Scale-location shows that there is perhaps some slight increase in the magnitude of standardized residuals as the fitted value increases but not to an extent that is a clear violation of assumptions.

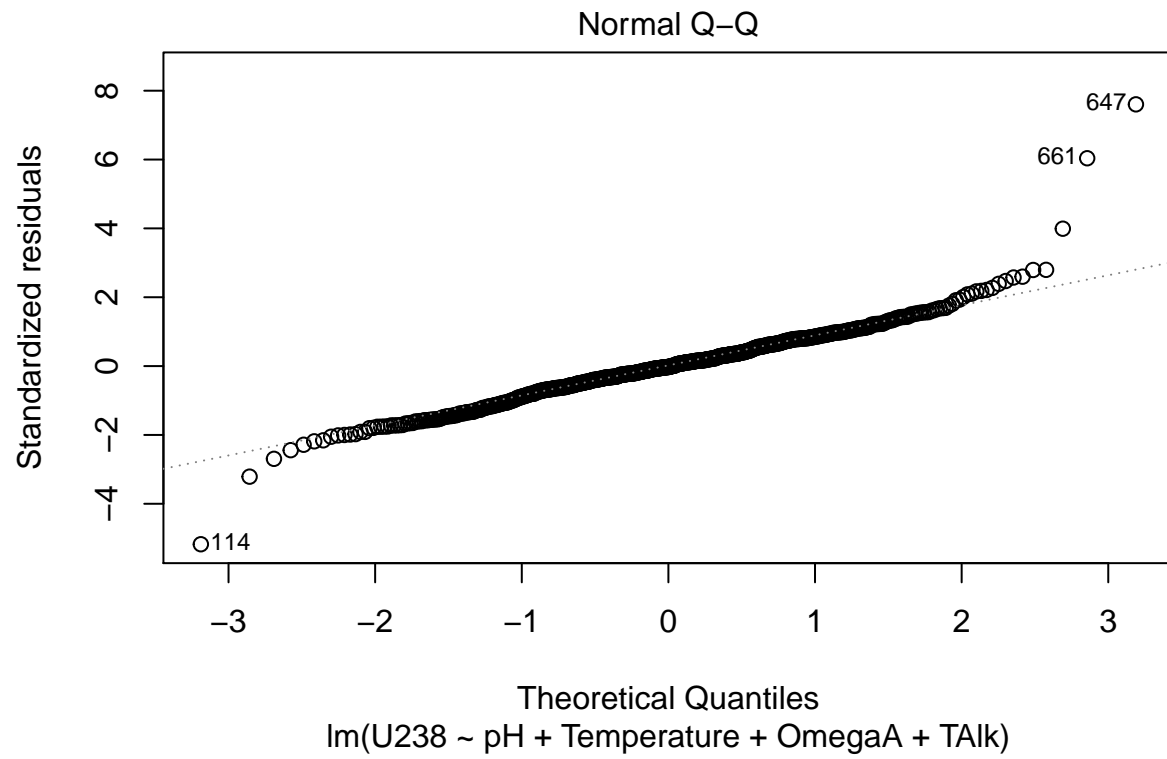
There are no concerning observations when looking at cook's distance in our residuals vs leverage plot.

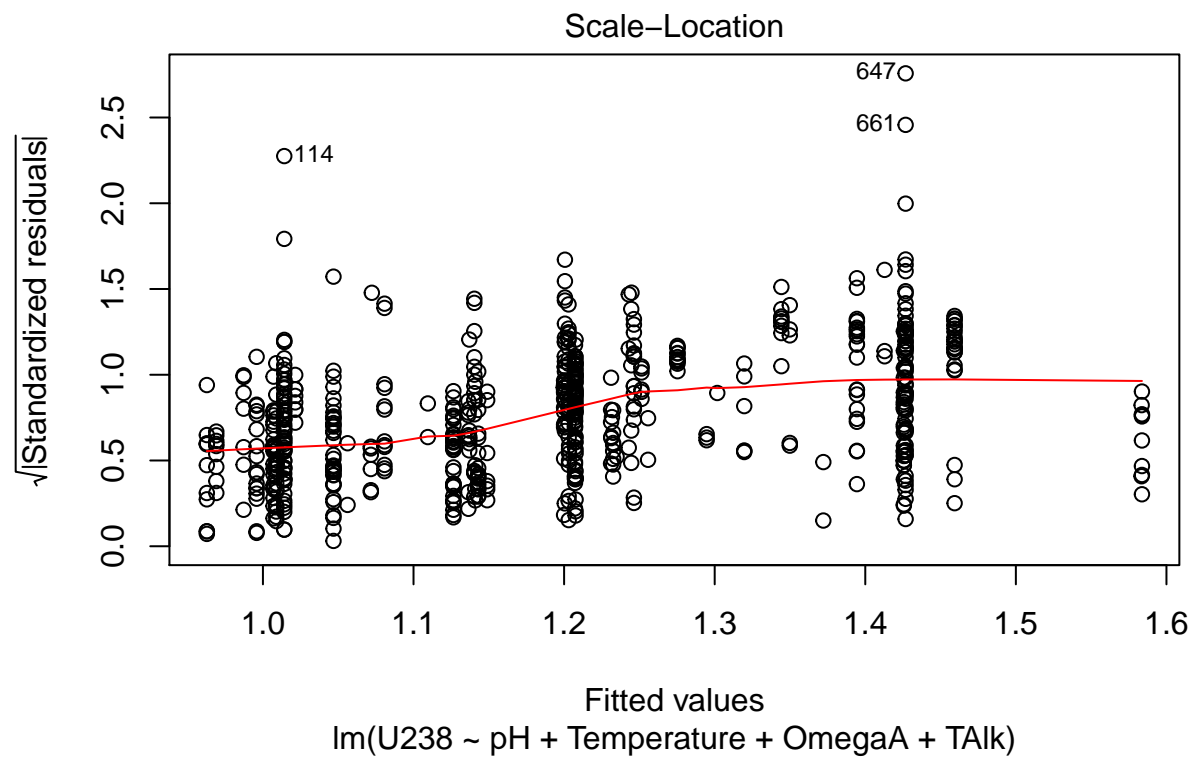
```
options(scipen=999)
final_model <- lm(U238 ~ pH + Temperature + OmegaA + TAlk, data=coral.df)
summary(final_model)
```

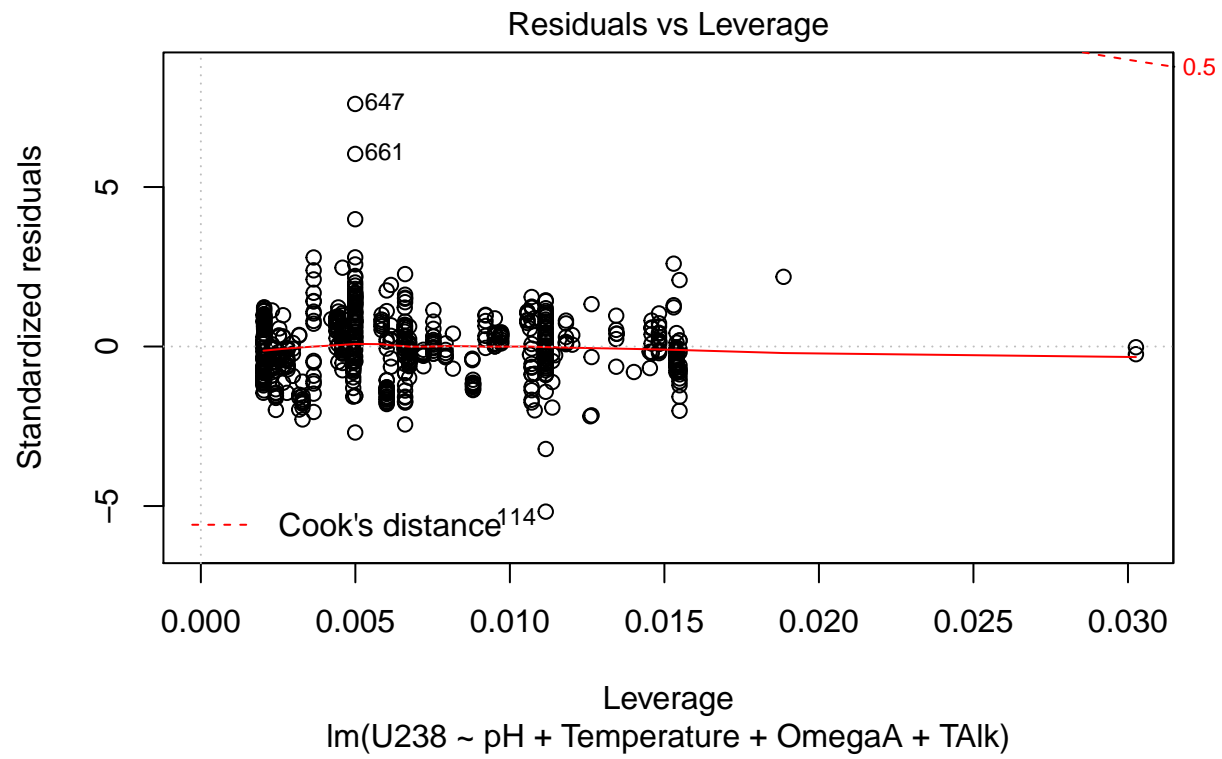
```
##
## Call:
## lm(formula = U238 ~ pH + Temperature + OmegaA + TAlk, data = coral.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90719 -0.09916 -0.00437  0.10731  1.33697
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -84.5117096  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 0.000000007673312738 ***
## ---
## 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)
```



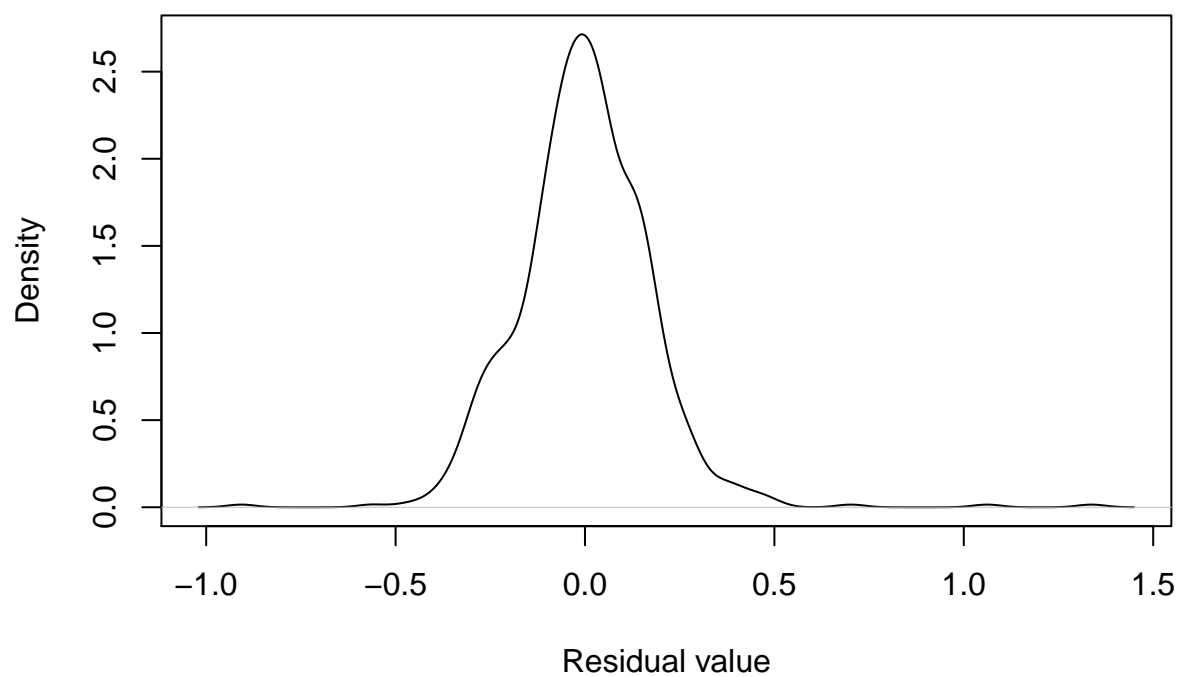






```
#normal Q-Q additional analysis
par(mfrow=c(1,1))
d<-density(final_model[['residuals']])
plot(d,main='Residual KDE Plot',xlab='Residual value')
```

Residual KDE Plot



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

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