

# 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", "Favia")
#remove coral with age > 10
coral <- coral[which(coral$Genus %in% genus_trim),] %>% dplyr::filter(Age < 10)

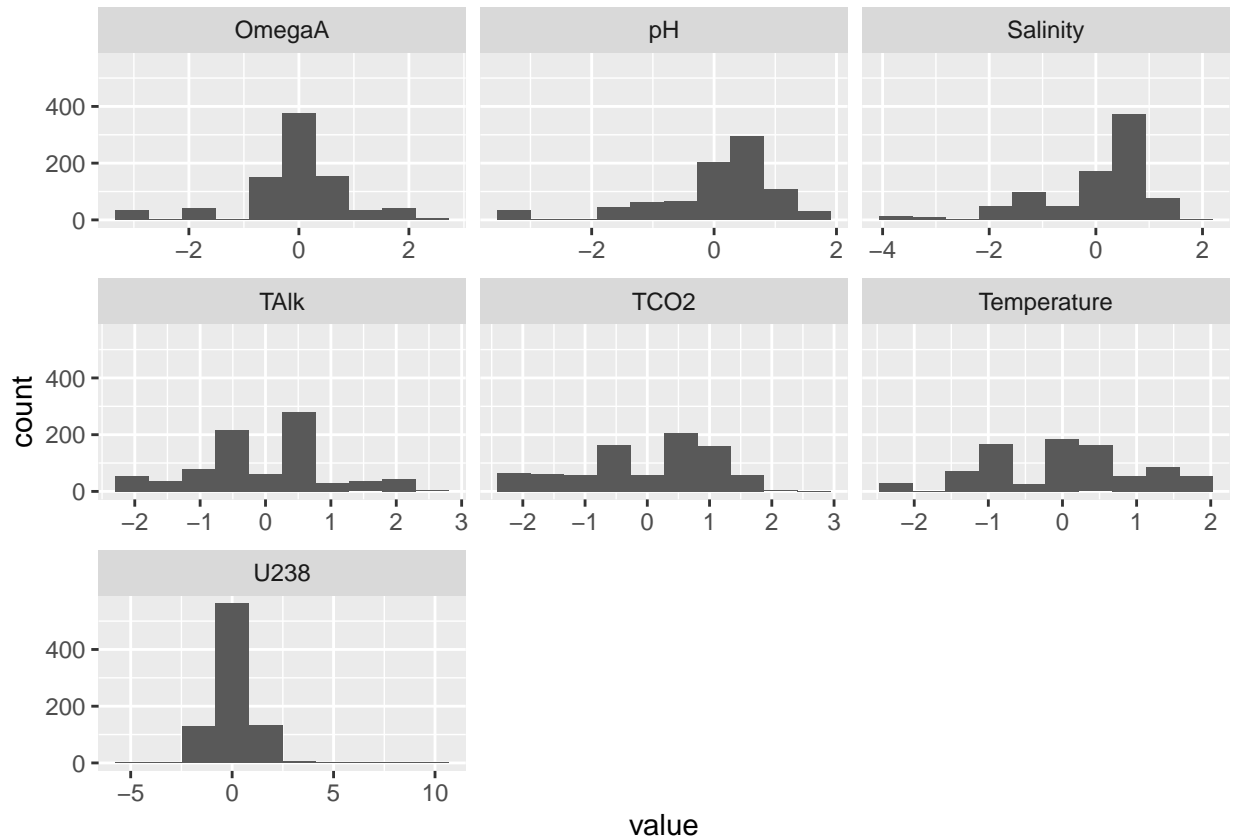
#clean up a nice dataframe
coral.df <- coral %>% mutate(Temperature = Temp) %>%
  select(U238,pH,TAlk,Salinity,Temperature,OmegaA,TCO2) %>%
  scale() %>%
  data.frame()
nrow(coral.df)
```

```
## [1] 835
```

## 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

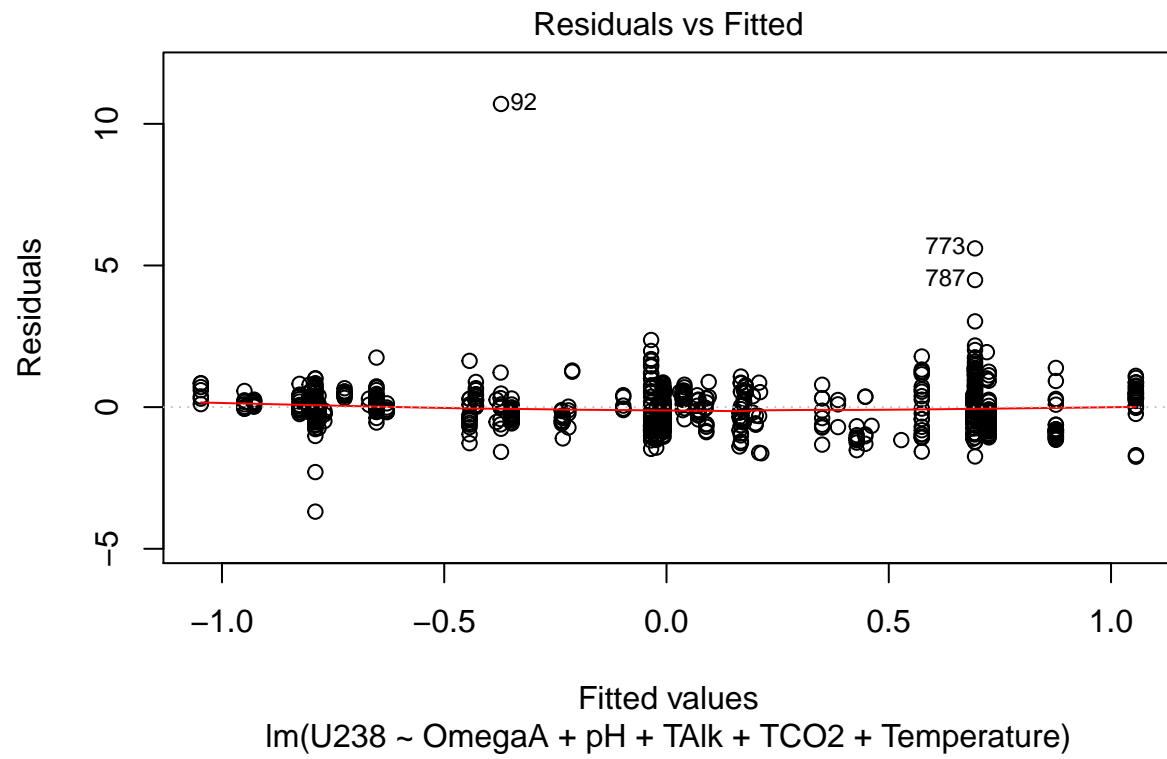
Residuals vs fitted looks perfectly normal - no signs of non-linearity or heteroskedasticity.

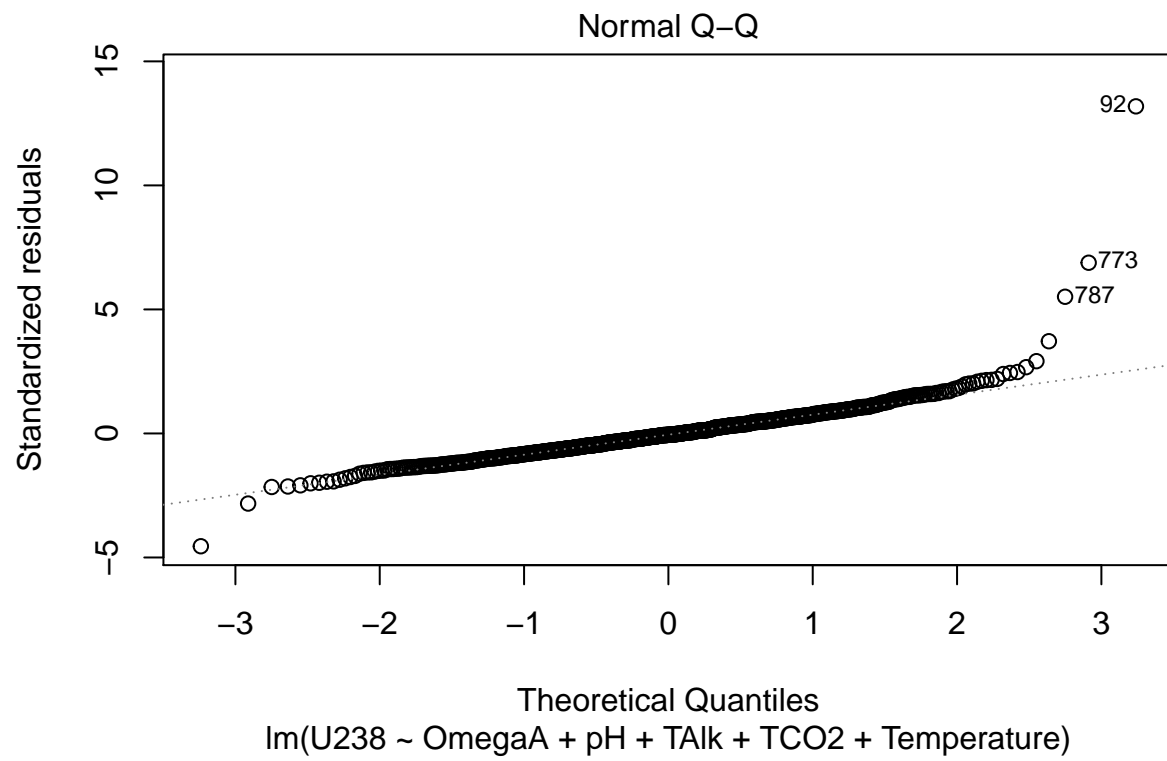
Normal Q-Q is somewhat concerning. According to the Shapiro test we reject the null hypothesis and conclude the data is non-normal.

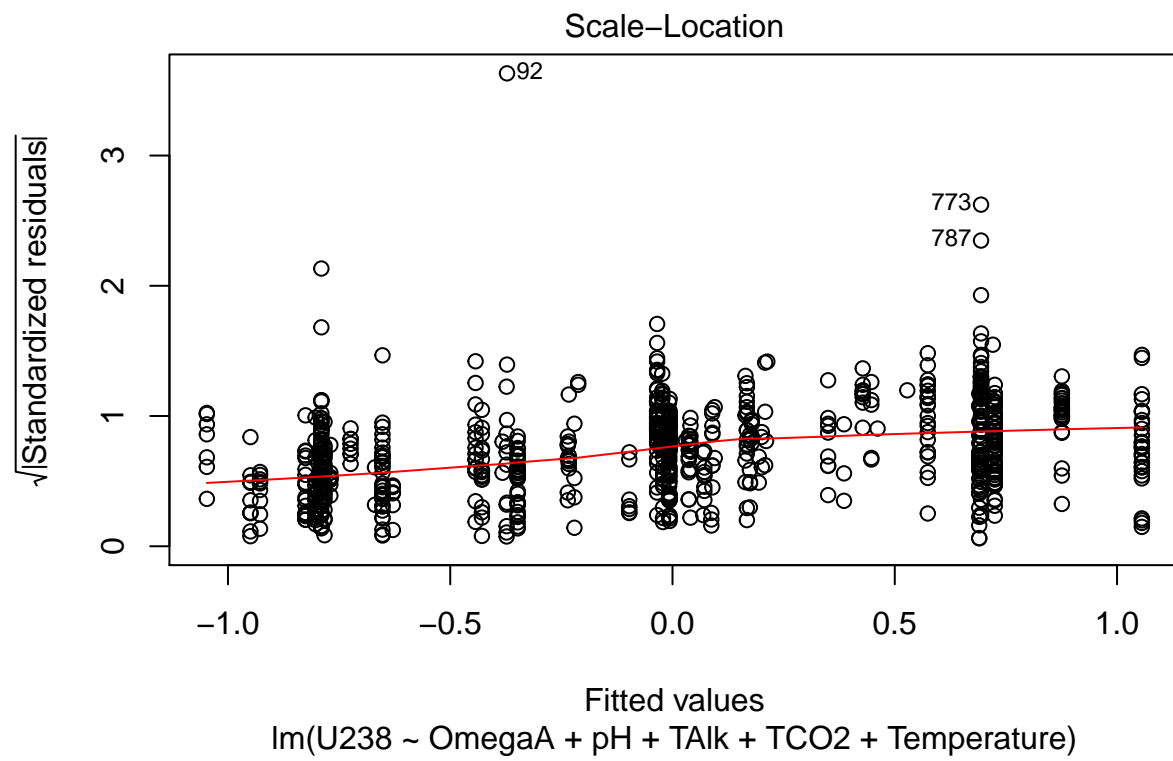
Scale-location is maybe slightly non-linear, but not to an extent that is cause for concern. We can also formally test heteroskedasticity with a Breusch-Pagan test. The null hypothesis is homoskedasticity, and we fail to reject it.

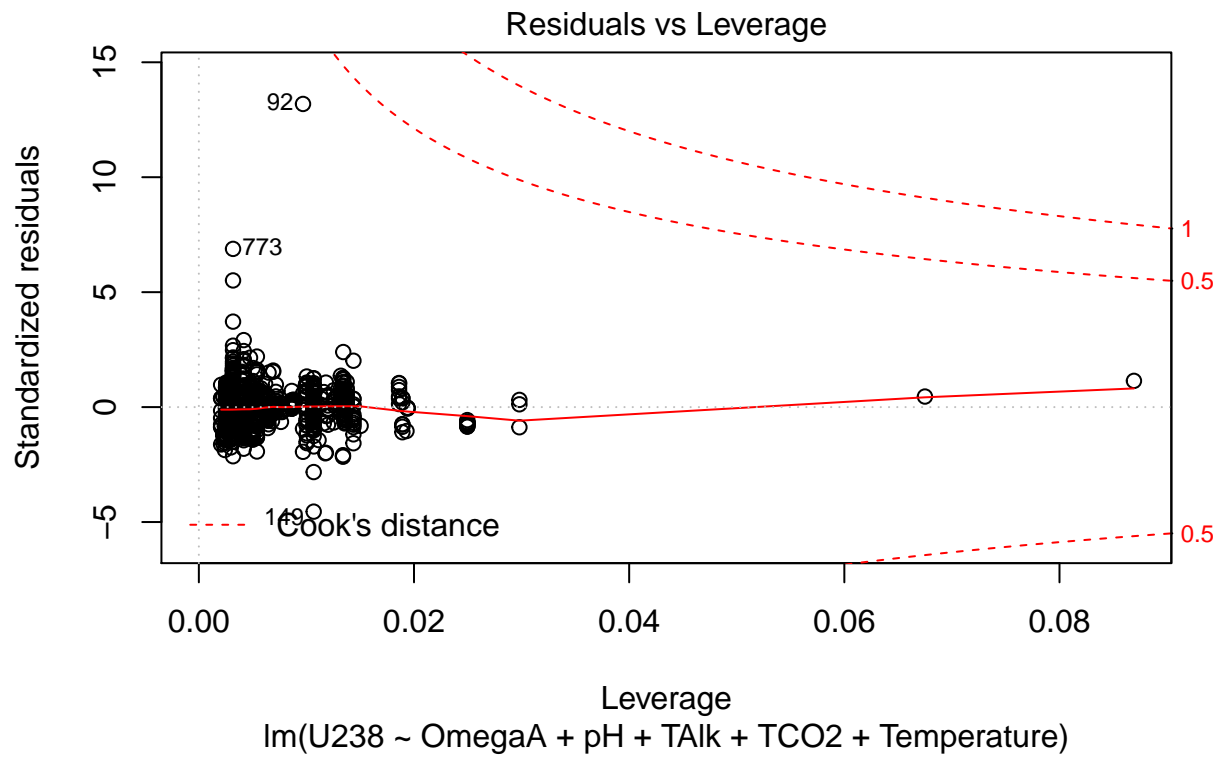
There are no concerning observations when looking at Cook's distance in our residuals vs leverage plot.

```
final_model <- lm(U238 ~ OmegaA + pH + TAlk + TCO2 + Temperature, data=coral.df)
plot(final_model)
```



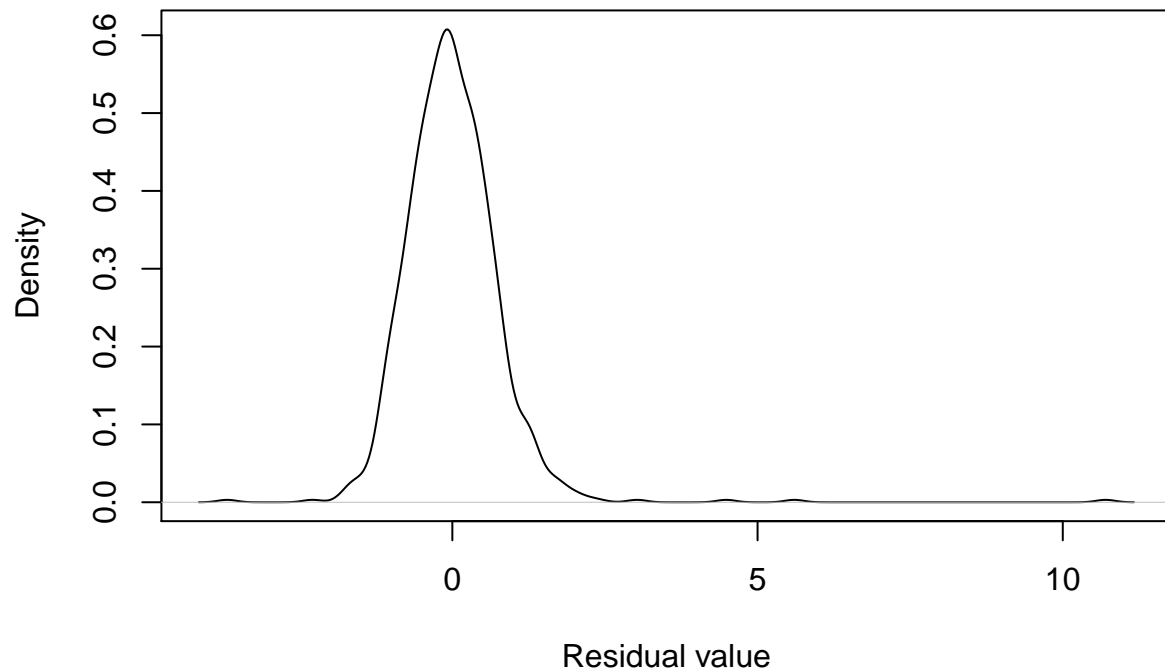






```
#normal Q-Q additional analysis
par(mfrow=c(1,1))
d<-density(final_model[['residuals']])
plot(d,main='Residual KDE Plot',xlab='Residual value')
```

## Residual KDE Plot



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

```
##  
## Shapiro-Wilk normality test  
##  
## data:  final_model[["residuals"]]  
## W = 0.82884, p-value < 2.2e-16
```

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

```
##  
## studentized Breusch-Pagan test  
##  
## data:  final_model  
## BP = 6.2438, df = 5, p-value = 0.2832
```