

220913 Linear Regression Model

```
library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

fish_df = read_csv("Fish.csv") %>%
  select(Species, Length3, Weight, Height, Width)

## Rows: 159 Columns: 7

## — Column specification


---


## Delimiter: ","
## chr (1): Species
## dbl (6): Weight, Length1, Length2, Length3, Height, Width
##
## ⓘ Use `spec()` to retrieve the full column specification for this data.
## ⓘ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

fish_model = lm(Width ~ Species + Length3 + Height + Weight, data = fish_df)

str(fish_df)

## tibble [159 × 5] (S3: tbl_df/tbl/data.frame)
##  $ Species: chr [1:159] "Bream" "Bream" "Bream" "Bream" ...
##  $ Length3: num [1:159] 30 31.2 31.1 33.5 34 34.7 34.5 35 35.1 36.2 ...
##  $ Weight  : num [1:159] 242 290 340 363 430 450 500 390 450 500 ...
##  $ Height  : num [1:159] 11.5 12.5 12.4 12.7 12.4 ...
##  $ Width   : num [1:159] 4.02 4.31 4.7 4.46 5.13 ...

factor(fish_df$Species)

##   [1] Bream    Bream    Bream    Bream    Bream    Bream    Bream
##   [8] Bream    Bream    Bream    Bream    Bream    Bream    Bream
##  [15] Bream    Bream    Bream    Bream    Bream    Bream    Bream
##  [22] Bream    Bream    Bream    Bream    Bream    Bream    Bream
##  [29] Bream    Bream    Bream    Bream    Bream    Bream    Bream
##  [36] Roach    Roach    Roach    Roach    Roach    Roach    Roach
```

```

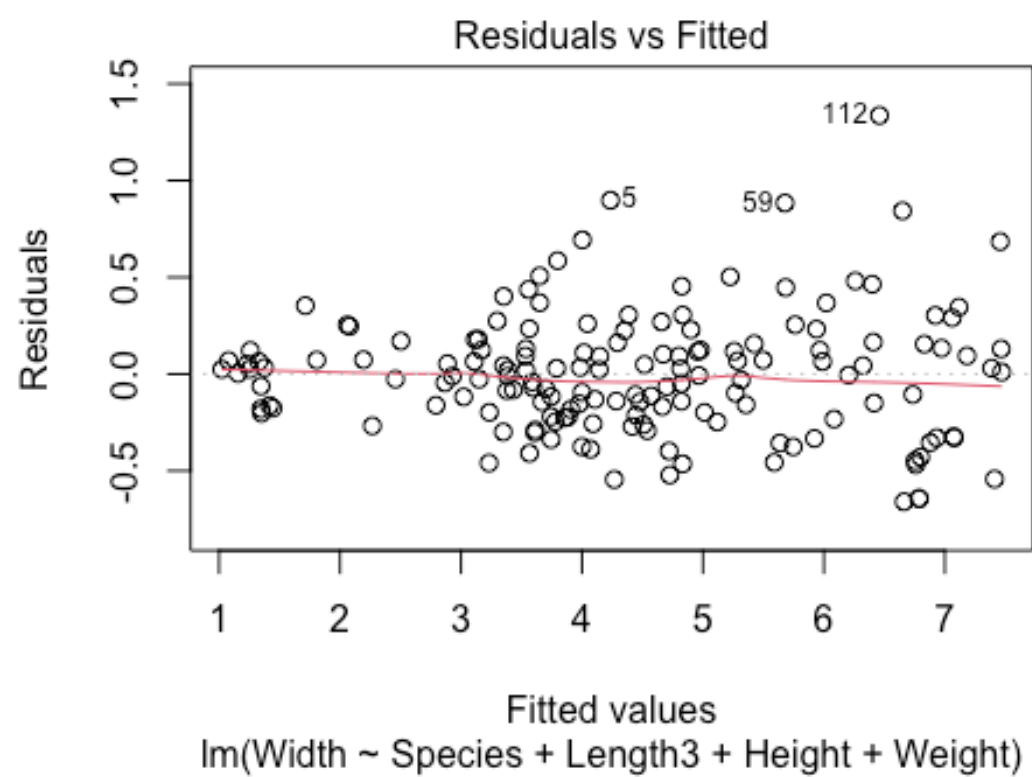
## [43] Roach      Roach      Roach      Roach      Roach      Roach      Roach
## [50] Roach      Roach      Roach      Roach      Roach      Roach      Roach
Whitefish
## [57] Whitefish Whitefish Whitefish Whitefish Whitefish Parkki   Parkki
## [64] Parkki     Parkki     Parkki     Parkki     Parkki     Parkki   Parkki
## [71] Parkki     Parkki     Perch      Perch      Perch      Perch    Perch
## [78] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [85] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [92] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [99] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [106] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [113] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [120] Perch      Perch      Perch      Perch      Perch      Perch    Perch
## [127] Perch      Perch      Pike        Pike        Pike        Pike     Pike
## [134] Pike        Pike        Pike        Pike        Pike        Pike     Pike
## [141] Pike        Pike        Pike        Pike        Pike        Pike     Smelt
## [148] Smelt       Smelt       Smelt       Smelt       Smelt       Smelt    Smelt
## [155] Smelt       Smelt       Smelt       Smelt       Smelt
## Levels: Bream Parkki Perch Pike Roach Smelt Whitefish

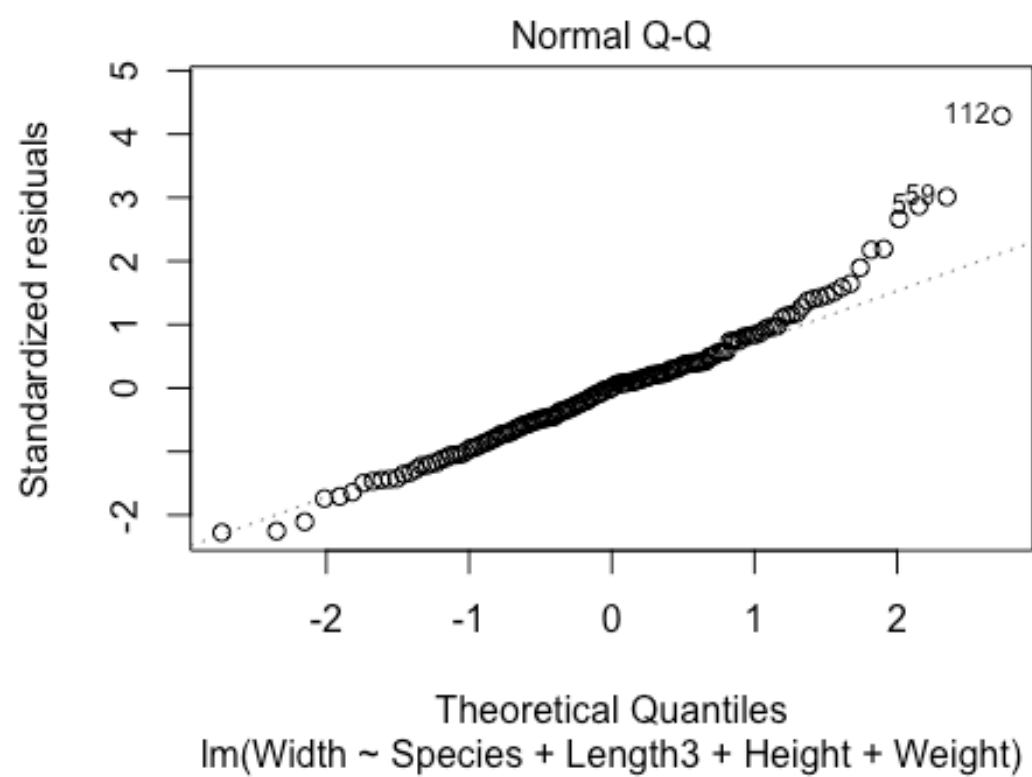
summary(fish_model)

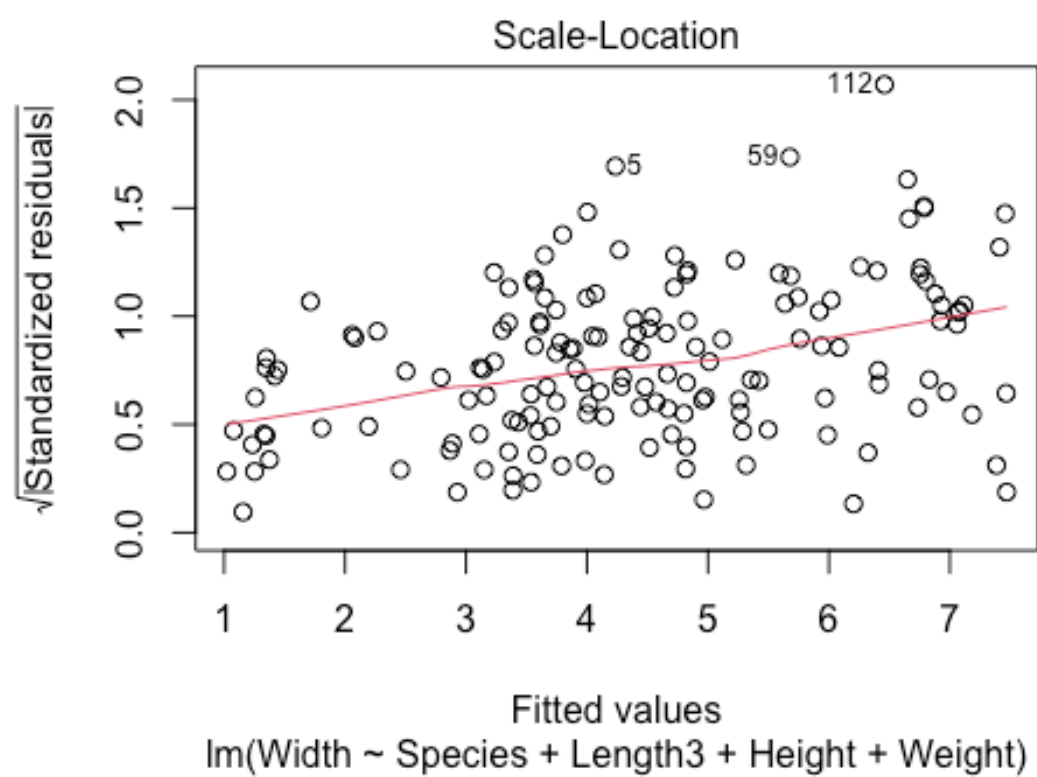
##
## Call:
## lm(formula = Width ~ Species + Length3 + Height + Weight, data = fish_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65959 -0.20046  0.00281  0.14401  1.33617
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.260e+00  3.263e-01  -6.926 1.21e-10 ***
## SpeciesParkki  9.212e-01  1.471e-01   6.263 3.84e-09 ***
## SpeciesPerch   2.303e+00  1.946e-01  11.834 < 2e-16 ***
## SpeciesPike    1.279e+00  3.933e-01   3.252 0.00142 **
## SpeciesRoach   1.921e+00  2.002e-01   9.594 < 2e-16 ***
## SpeciesSmelt   1.921e+00  2.752e-01   6.980 9.04e-11 ***
## SpeciesWhitefish 1.991e+00  2.058e-01   9.677 < 2e-16 ***
## Length3        7.465e-02  1.473e-02   5.069 1.17e-06 ***
## Height         3.198e-01  3.615e-02   8.848 2.40e-15 ***
## Weight        -5.012e-05  2.793e-04  -0.179 0.85784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.323 on 149 degrees of freedom
## Multiple R-squared:  0.9654, Adjusted R-squared:  0.9633
## F-statistic: 461.6 on 9 and 149 DF,  p-value: < 2.2e-16

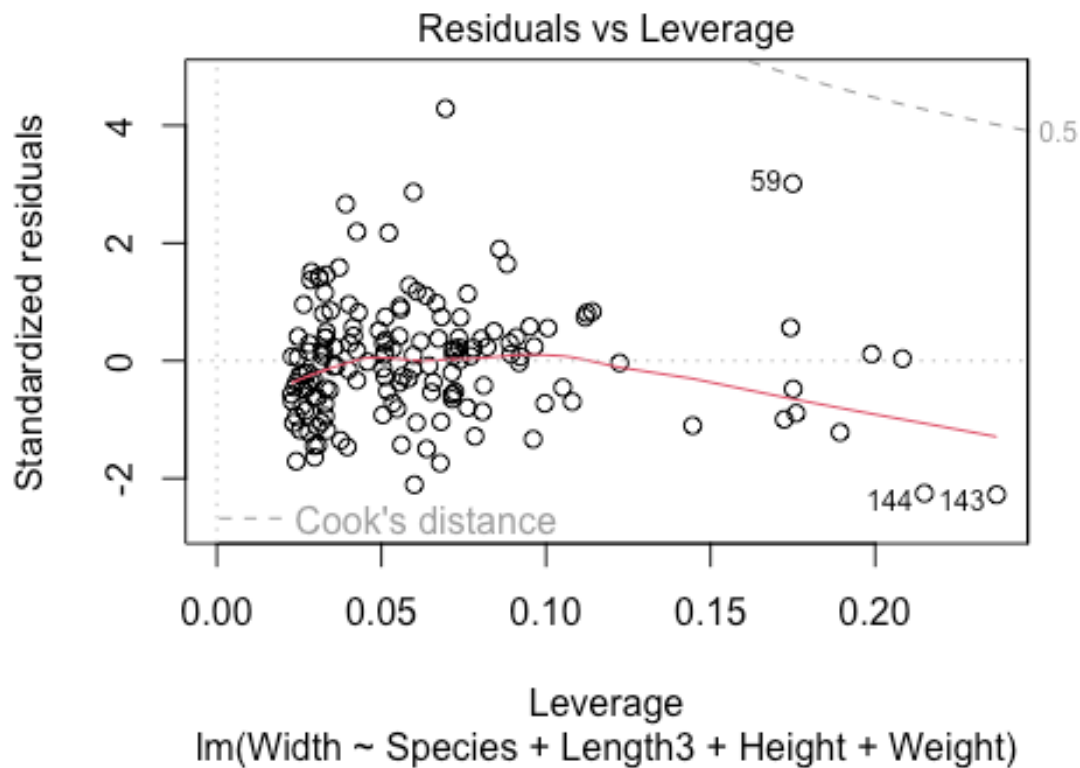
```

```
plot(fish_model) # 첫번째 plot을 보면 Residuals vs Fitted 가 heterogeneous. >  
Homoscedaticity 만족 x.
```









Normal Q-Q > 정균분포를 따른다고 가정하므로, 0쪽에 밀집, 양쪽으로 갈수록 퍼져야 함. 오른쪽 부분을 보면 맞지 않다.

Scale-Location: X-Y 가 어떠한 경향인지 확인하는. 빨간 선은 수평이 되어야 함. 그리고 고르게 퍼져있어야 한다. 좋지 않음. Homoscedaticity 만족 X

Residuals vs Leverage: Outlier에 의해서 얼마나 쏠리는지 (Outlier가 있을 때와 없을 때.) Cook's distance가 일정(수평선)해야 함. 절대값이 1을 넘으면 안된다. 아무리 커도 3을 넘으면 안된다.

#independence 확인

plot(fish_df)

