More data analysis

Derek Situ

2022-03-11

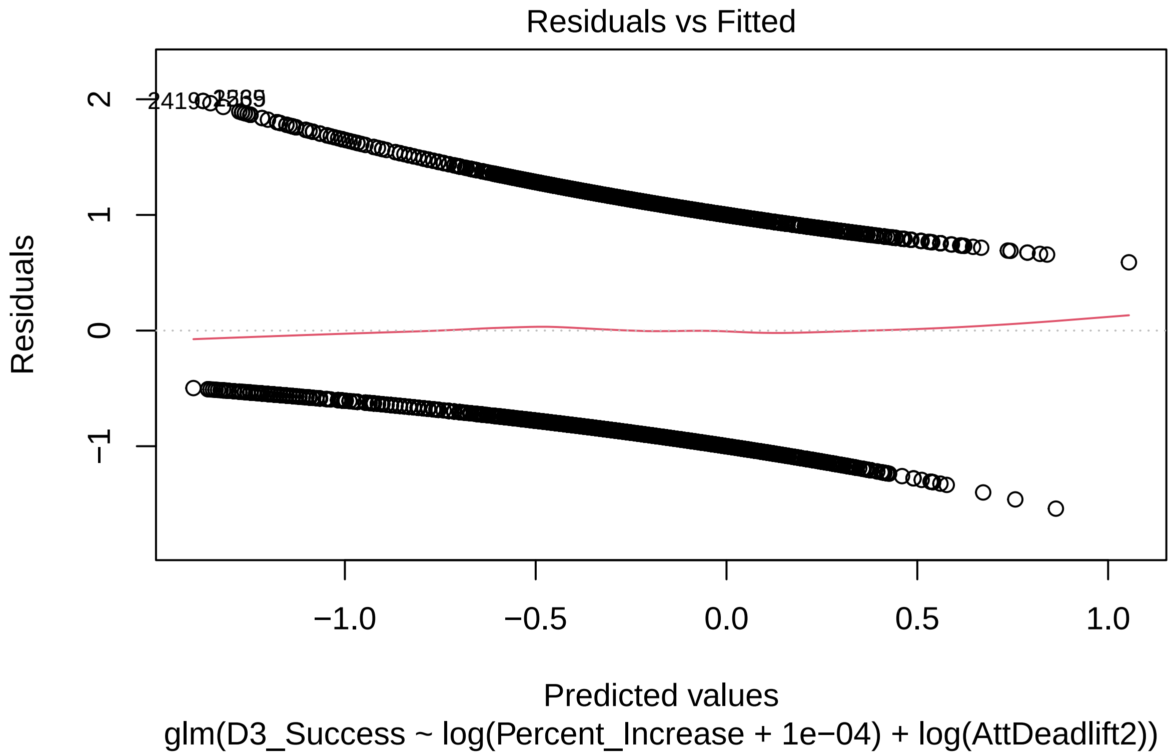
Reading in the data

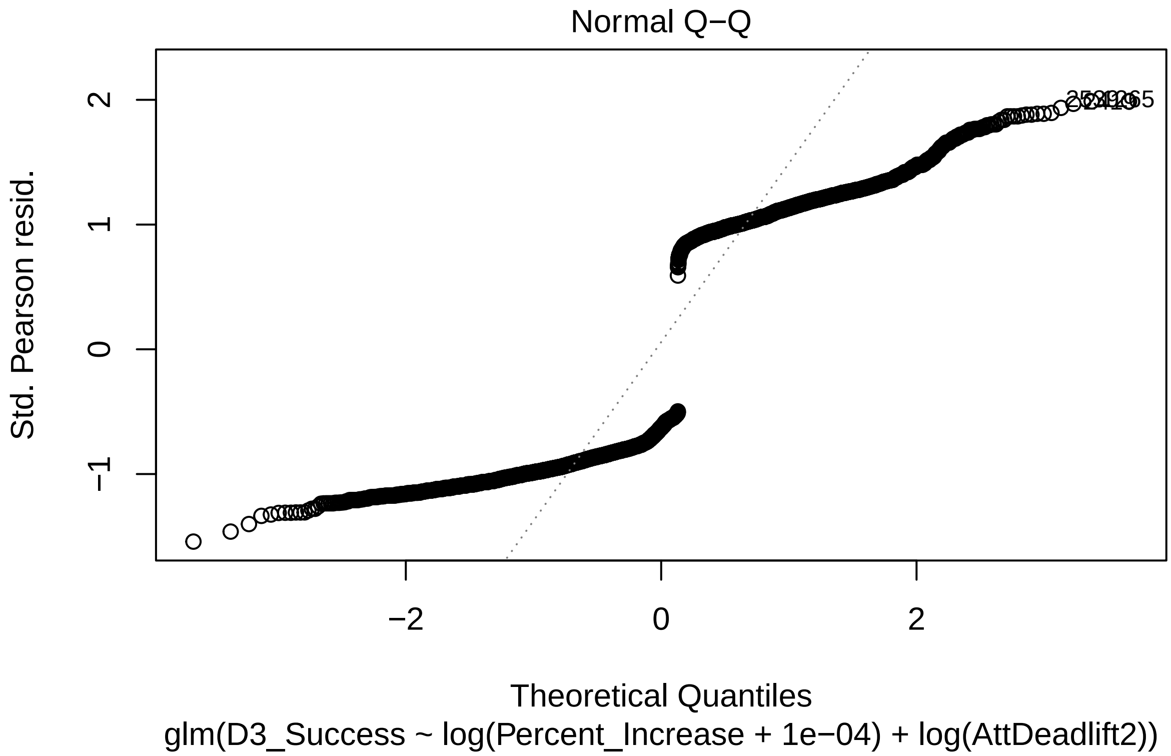
```
# read in a csv with file path "path", keep only necessary columns and rename
#   for readability, add column "Subtotal" which is equal to the sum of
#   BestSquat, BestBench, and max(0, Deadlift1, Deadlift2)
read_results <- function(path) {
  read_csv(path, col_names = TRUE) %>%
    select(Name, Sex, Division,
           WeightClass = "WeightClassKg", Bodyweight = "BodyweightKg",
           Squat1 = "Squat1Kg", Squat2 = "Squat2Kg", Squat3 = "Squat3Kg",
           BestSquat = "Best3SquatKg",
           Bench1 = "Bench1Kg", Bench2 = "Bench2Kg", Bench3 = "Bench3Kg",
           BestBench = "Best3BenchKg",
           Deadlift1 = "Deadlift1Kg", Deadlift2 = "Deadlift2Kg",
           Deadlift3 = "Deadlift3Kg", BestDeadlift = "Best3DeadliftKg",
           Total = "TotalKg", Place) %>%
    mutate(Subtotal = BestSquat + BestBench +
             ifelse(Deadlift1 > 0 | Deadlift2 > 0,
                    ifelse(Deadlift2 > Deadlift1, Deadlift2, Deadlift1),
                    0)) %>%
    relocate(Subtotal, .before = Deadlift3)
}

# read in data
data_21 <- read_results("2021.csv")
data_19 <- read_results("2019.csv")
data_18 <- read_results("2018.csv")
data_17 <- read_results("2017.csv")
data_16 <- read_results("2016.csv")
data_15 <- read_results("2015.csv")
data_14 <- read_results("2014.csv")
data_13 <- read_results("2013.csv")
data_12 <- read_results("2012.csv")

# merge all datasets so that we can run regressions with all of the data
  # note that when a weight n is attempted and failed, it is entered as -n.
  # that's why we calculate attempted lifts as the absolute value of the entry.
all_data <- rbind(data_21, data_19, data_18, data_17, data_16, data_15,
                  data_14, data_13, data_12) %>%
  select(Name, WeightClass, Bodyweight, Squat3, Deadlift2, Deadlift3) %>%
  mutate(AttSquat3 = abs(Squat3), # Att is short for attempted.
         AttDeadlift2 = abs(Deadlift2),
         AttDeadlift3 = abs(Deadlift3),
         Percent_Increase = ((AttDeadlift3 - AttDeadlift2) / AttDeadlift2) * 100,
         D3_Success = ifelse(Deadlift3 > 0, 1, 0)) %>% # whether Deadlift3 is a success.
  filter(!(WeightClass %in% c("84+", "105", "120", "120+"))) # so that we get a linear fit
```

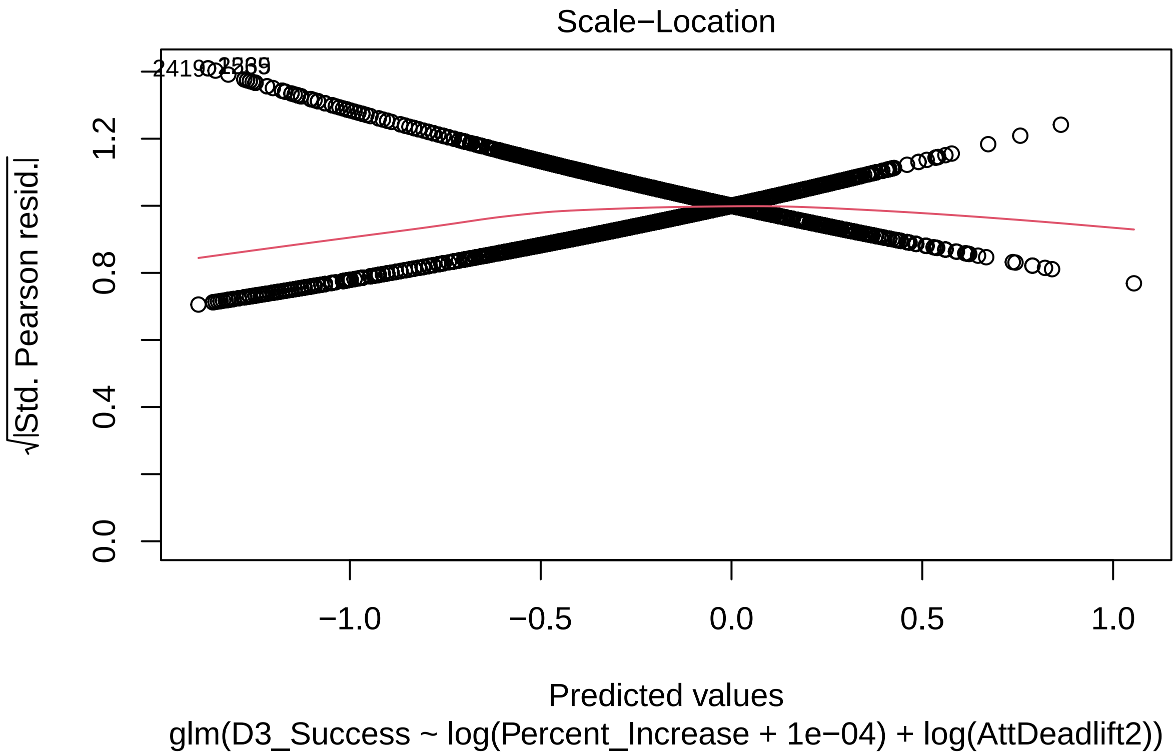
```
lm_fit <- glm(D3_Success ~ log(Percent_Increase + 0.0001) + log(AttDeadlift2),
              family = binomial(link = "logit"),
              data = all_data)
summary(lm_fit)
```

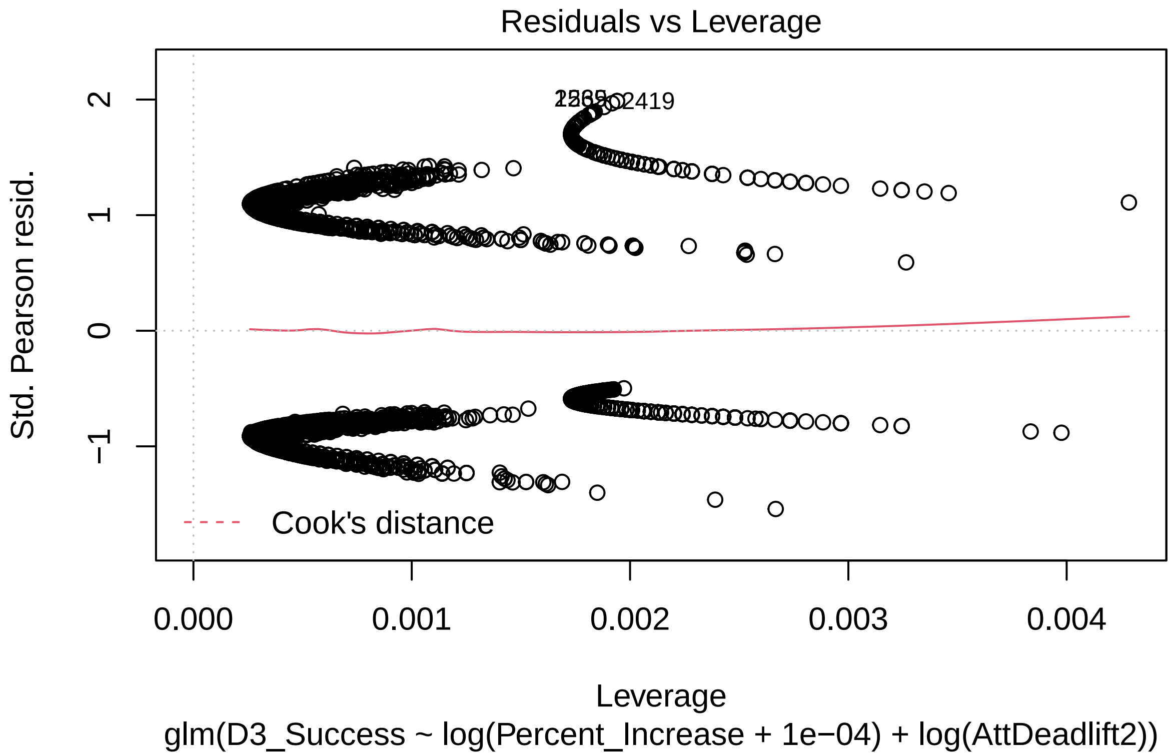
```
## 
## Call:
## glm(formula = D3_Success ~ log(Percent_Increase + 1e-04) + log(AttDeadlift2), 
##     family = binomial(link = "logit"), data = all_data)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -1.5589  -1.0969  -0.8175   1.1976   1.7876  
## 
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                    4.130379   0.556710   7.419 1.18e-13 ***
## log(Percent_Increase + 1e-04)  0.074240   0.009453   7.854 4.04e-15 ***
## log(AttDeadlift2)             -0.833056   0.106810  -7.799 6.22e-15 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 5530.5  on 4020  degrees of freedom
## Residual deviance: 5392.0  on 4018  degrees of freedom
##   (123 observations deleted due to missingness)
## AIC: 5398
## 
## Number of Fisher Scoring iterations: 4
```

```
plot(lm_fit)
```

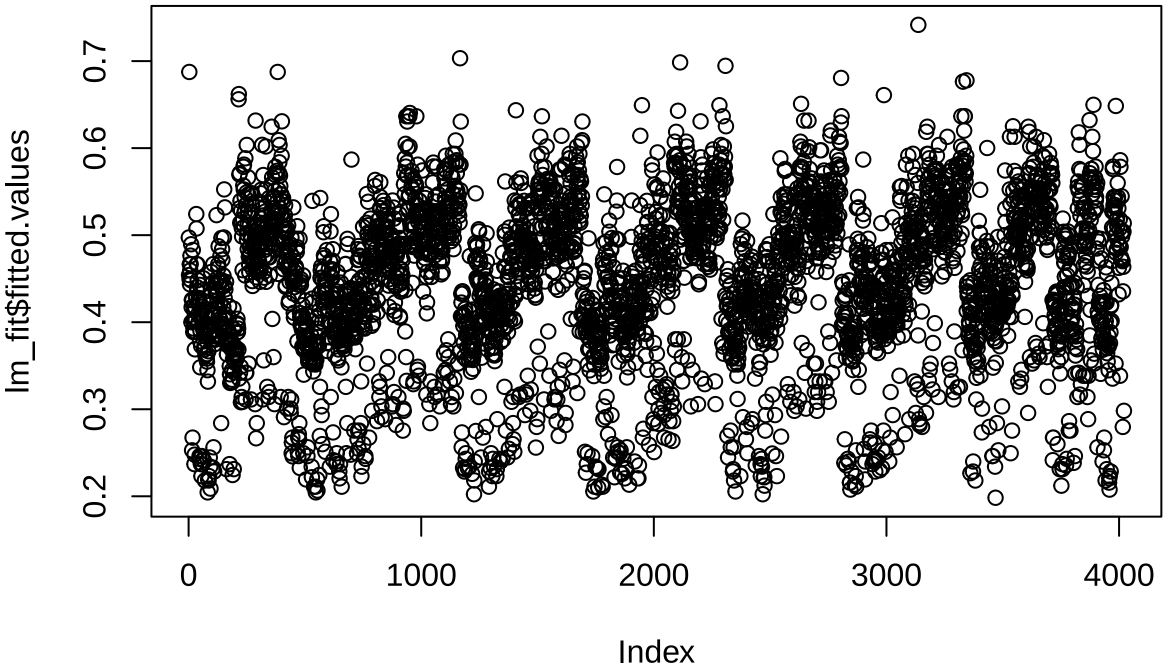

Residuals vs Fitted
Residuals
2
1
0
−1
2419
2565
−1.0
−0.5
0.0
0.5
1.0
Predicted values
glm(D3_Success ~ log(Percent_Increase + 1e−04) + log(AttDeadlift2))


Normal Q−Q
Std. Pearson resid.
2
1
0
−1
−2
0
2
Theoretical Quantiles
glm(D3_Success ~ log(Percent_Increase + 1e−04) + log(AttDeadlift2))

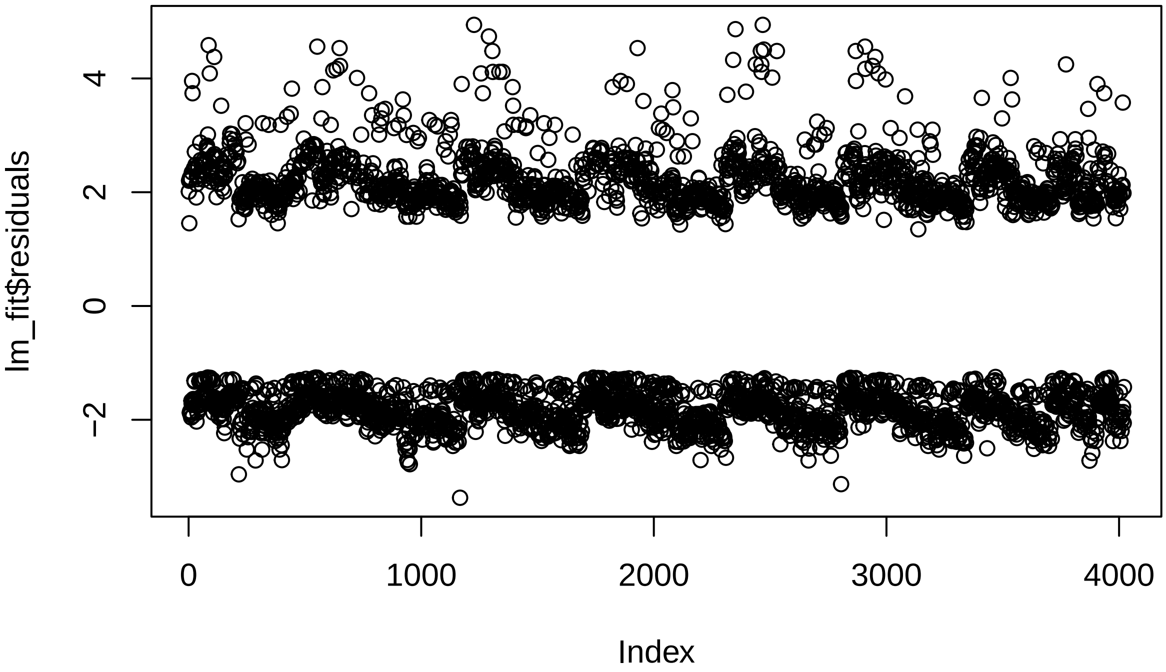

Scale−Location
2419
1569
2585
1.2
0.8
0.4
0.0
√|Std. Pearson resid.|
−1.0
−0.5
0.0
0.5
1.0
Predicted values
glm(D3_Success ~ log(Percent_Increase + 1e−04) + log(AttDeadlift2))



```
plot(lm_fit$fitted.values)
```



```
plot(lm_fit$residuals)
```



pretty terrible fit