Exploratory variables are geometry coefficients and the Froude number:

1. Longitudinal position of the center of buoyancy, adimensional.
2. Prismatic coefficient, adimensional.
3. Length-displacement ratio, adimensional.
4. Beam-draught ratio, adimensional.
5. Length-beam ratio, adimensional.
6. Froude number, adimensional.

The target histogram negative variable is the residuary resistance per unit weight of displacement:

7. Resistance: Residuary resistance per unit weight of displacement, adimensional.

Take the log of the target?

Yes, if we plot *Resistance* we see a significant right skew. However if we plot the histogram of the logarithm of *Resistance*, we see a distribution that looks much more like a normal distribution:

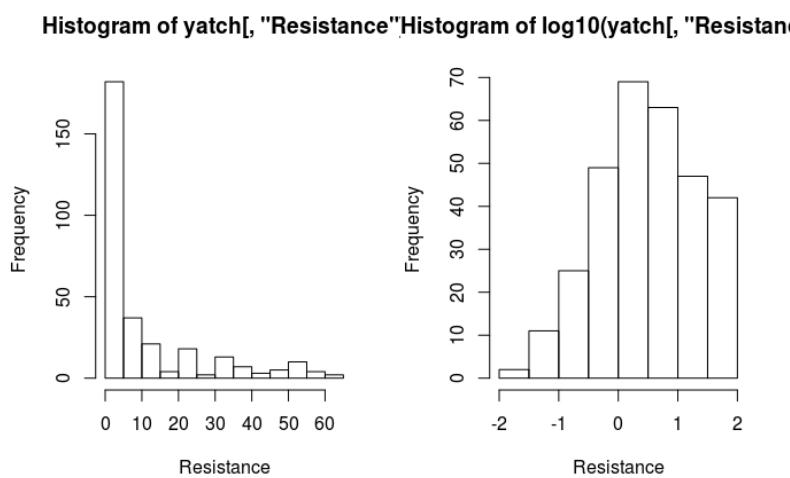

Figure 1 - log of the target

LINEAR REGRESSION

Only correlation *Froude.No – Resistance* seems significant, other columns are omited.

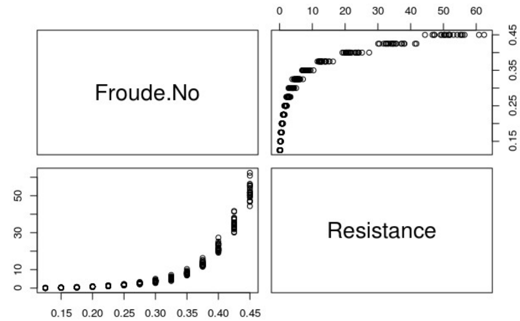

Figure 2 - correlations

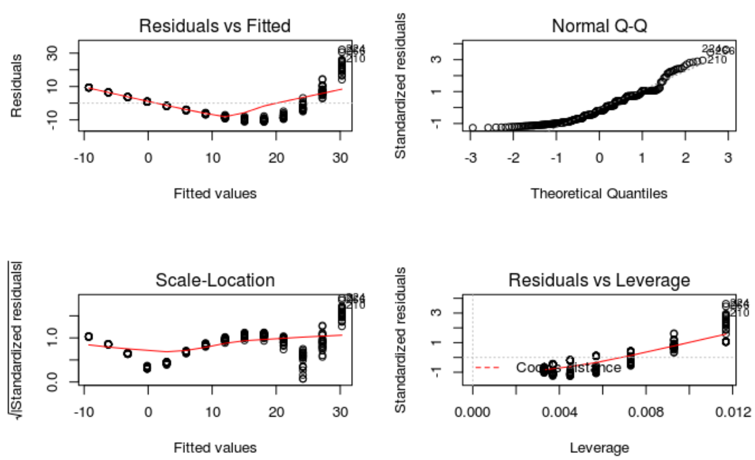

Figure 3 - model main plots

Independent variables, as well as dependent, are transformed:

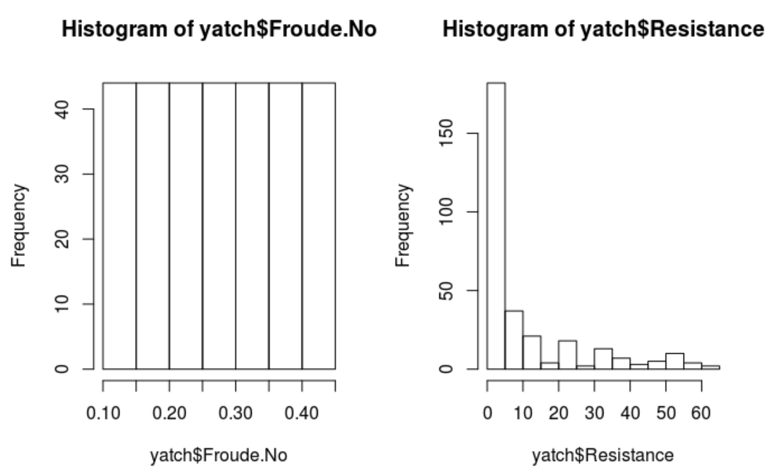

Figure 4 - original histogram independent/dependant var

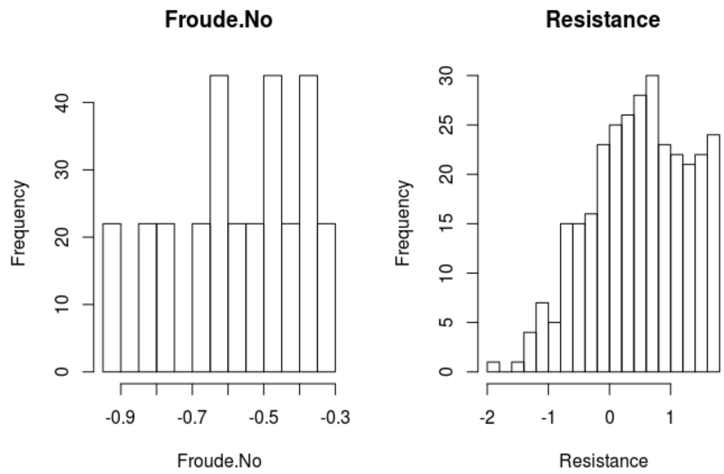

Figure 5 - histogram transformed variables - log10(x)

```
Call:
lm(formula = yacht$Resistance ~ ., data = yacht)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.240  -7.669  -1.726   6.404  32.154 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -24.484      1.534  -15.96   <2e-16 ***
Froude.No    121.668      5.034   24.17   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.903 on 306 degrees of freedom
Multiple R-squared:  0.6562, Adjusted R-squared:  0.6551 
F-statistic: 584.2 on 1 and 306 DF,  p-value: < 2.2e-16
```

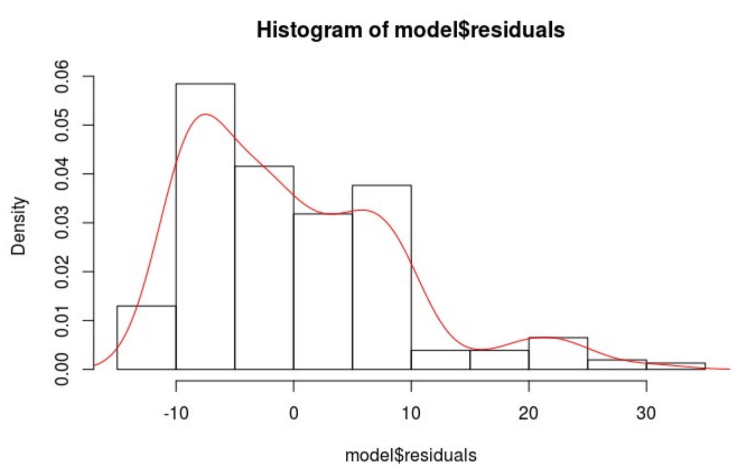

Figure 6 - Histogram of model residuals

Taking the whole variables does not provide any special benefit:

```
Call:
lm(formula = yacht$Resistance ~ ., data = yacht)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.770  -7.565  -1.881   6.112  31.572 

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                   -19.2367    27.1133  -0.709    0.479    
Longitudinal.pos                                0.1938     0.3381   0.573    0.567    
Prismatic.coeff                                -6.4194    44.1590  -0.145    0.885    
Length.displ                                    4.2330    14.1651   0.299    0.765    
Beam.draught..................................  -1.7657     5.5212  -0.320    0.749    
Length.beam                                    -4.5164    14.2000  -0.318    0.751    
Froude.No                                     121.6676     5.0658  24.018   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.96 on 301 degrees of freedom
Multiple R-squared:  0.6576,	Adjusted R-squared:  0.6507 
F-statistic: 96.33 on 6 and 301 DF,  p-value: < 2.2e-16
```

Our model does not look fine, even with centering, scaling, or with the already applied *log* transform, results do not improve much.

RIDGE REGRESSION

Find lambdas:

```
> lambdes <- seq(0.001,0.5,0.001)
> select(lm.ridge(model, lambda = lambdes))
```

Get model:

```
> lambdes[which.min(model.ridge$GCV)]
0.001
> model.ridge<-lm.ridge(model, lambda = 0.001)
```

Prediction error:

```
> norm.root.mse.LOOCV
> sqrt(model.ridge$GCV)
0.5147124

...
```