Hagfish length, weight, diameter relationships

Ben Williams, Jane Sullivan

January 24, 2019

Contents

Questions

- calculate the diameter of a hagfish based upon length and weight,
- calculate the hole size that a hagfish can squeeze through.

Exploratory data analysis

First a couple of basic examinations of the data.

Models

A suite of models were examined to explore diameter ~ length/weight relationships. Generalized additive models and generalized linear models with and without interaction and/or log links were examined.

```
gam_l <- gam(diameter ~ s(length, k=4), data = hag, gamma = 1.4, method='ML')
gam_w <- gam(diameter ~ s(weight, k=4), data = hag, gamma = 1.4, method='ML')
gam_lw <- gam(diameter ~ s(length, k=4) + s(weight, k=4), data = hag, gamma = 1.4, method='ML')
gam_llw <- gam(diameter ~ s(length, k=4) + s(weight, k=4), data = hag, gamma = 1.4, method='ML', family


glm_l <- glm(diameter ~ length, data = hag)
glm_w <- glm(diameter ~ weight, data = hag)
glm_lw <- glm(diameter ~ length * weight, data = hag)
glm_llw <- glm(diameter ~ length * weight, data = hag, family = gaussian(link = 'log'))

null_lm <- lm(diameter ~ length, data = hag)
```

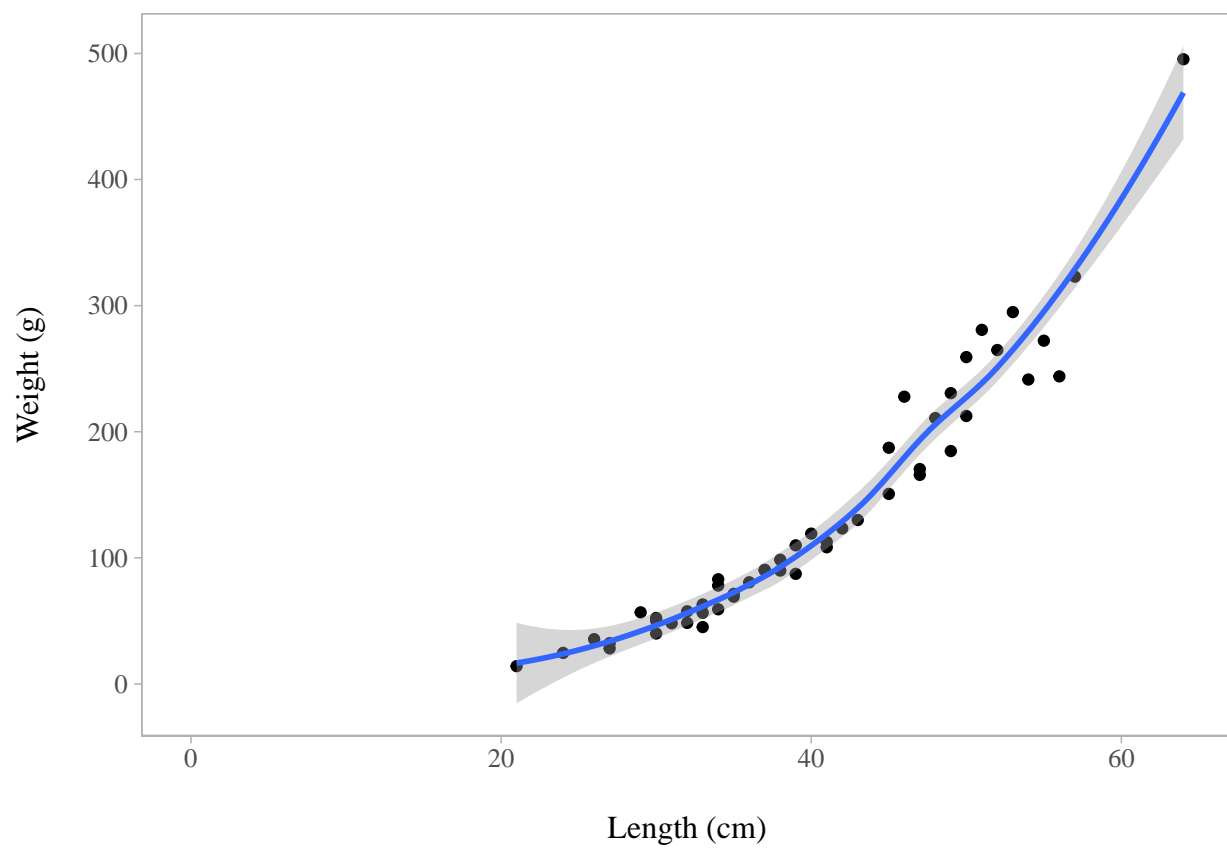AIC was used to test for the top model.
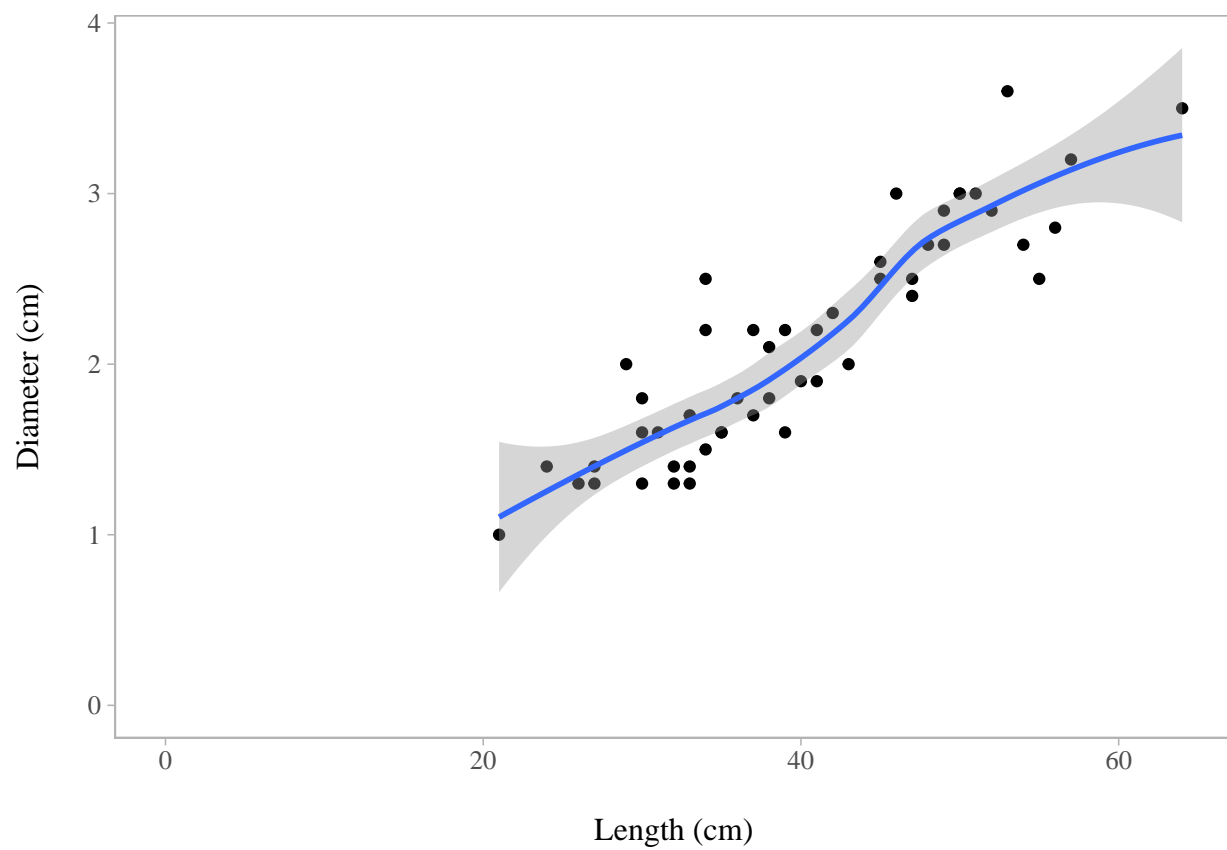
Figure 1: Hagfish length-weight measurements.

Figure 2: Hagfish length-diameter measurements.
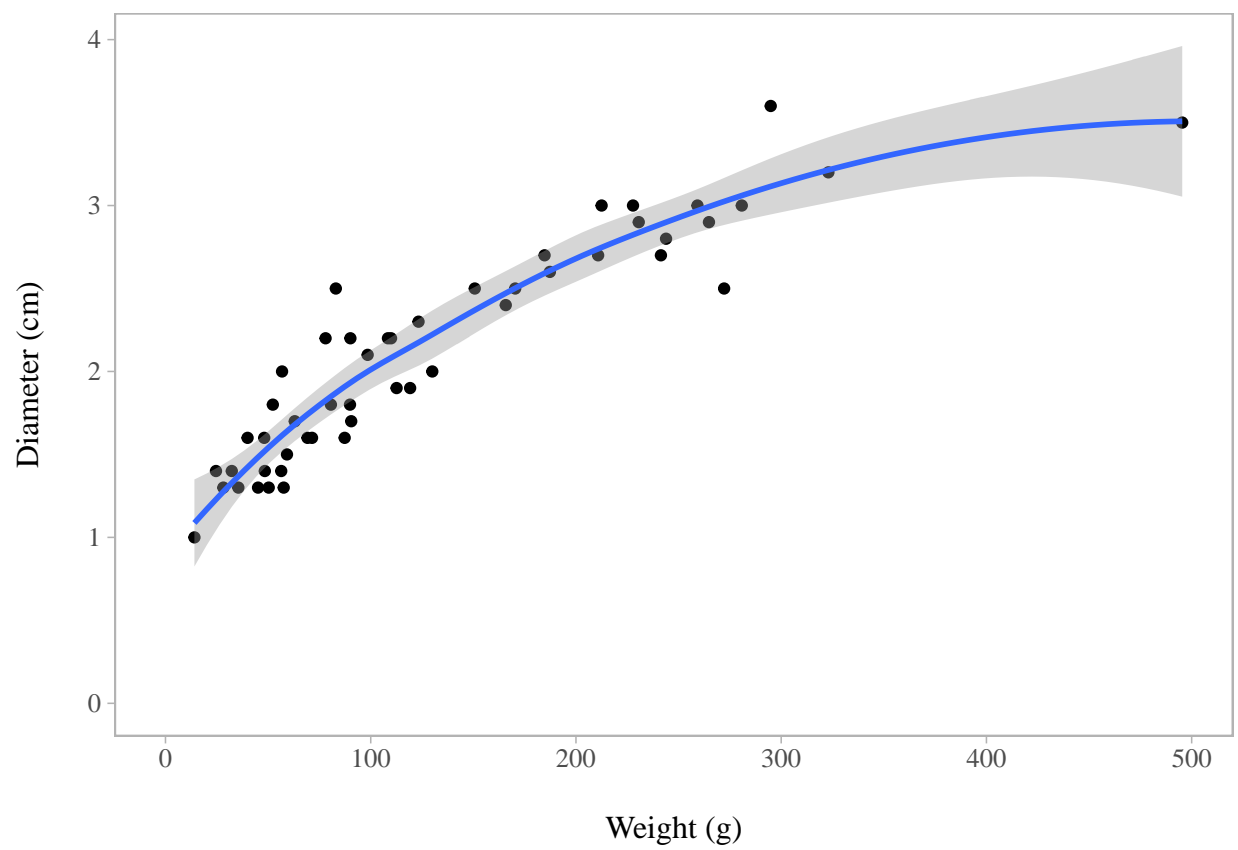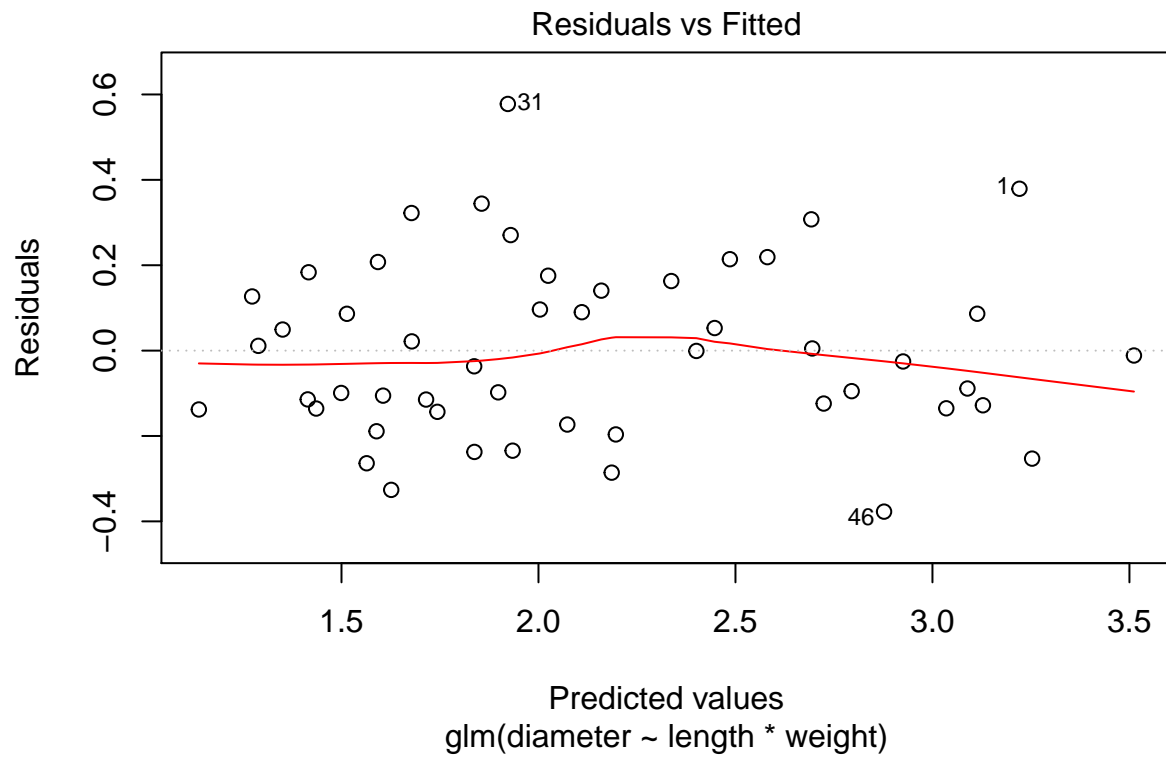
Figure 3: Hagfish weight-diameter measurements.
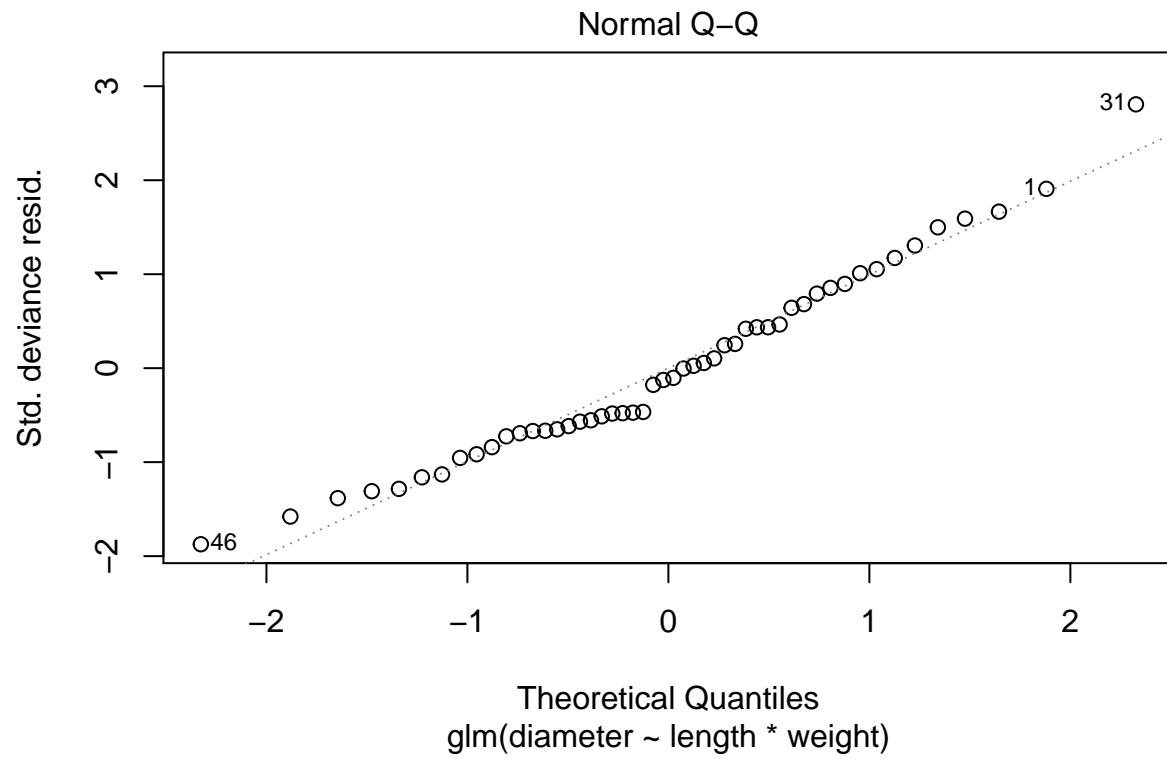
```
##                df        AIC
## gam_l    3.000005 21.210596
## gam_w    4.806820 -1.215236
## gam_lw   5.954371 -7.568013
## gam_llw  5.977033 -3.874416
## glm_l    3.000000 21.210589
## glm_w    3.000000 16.708175
## glm_lw   5.000000 -8.232758
## glm_llw  5.000000 -2.935581
## null_lm  3.000000 21.210589
```
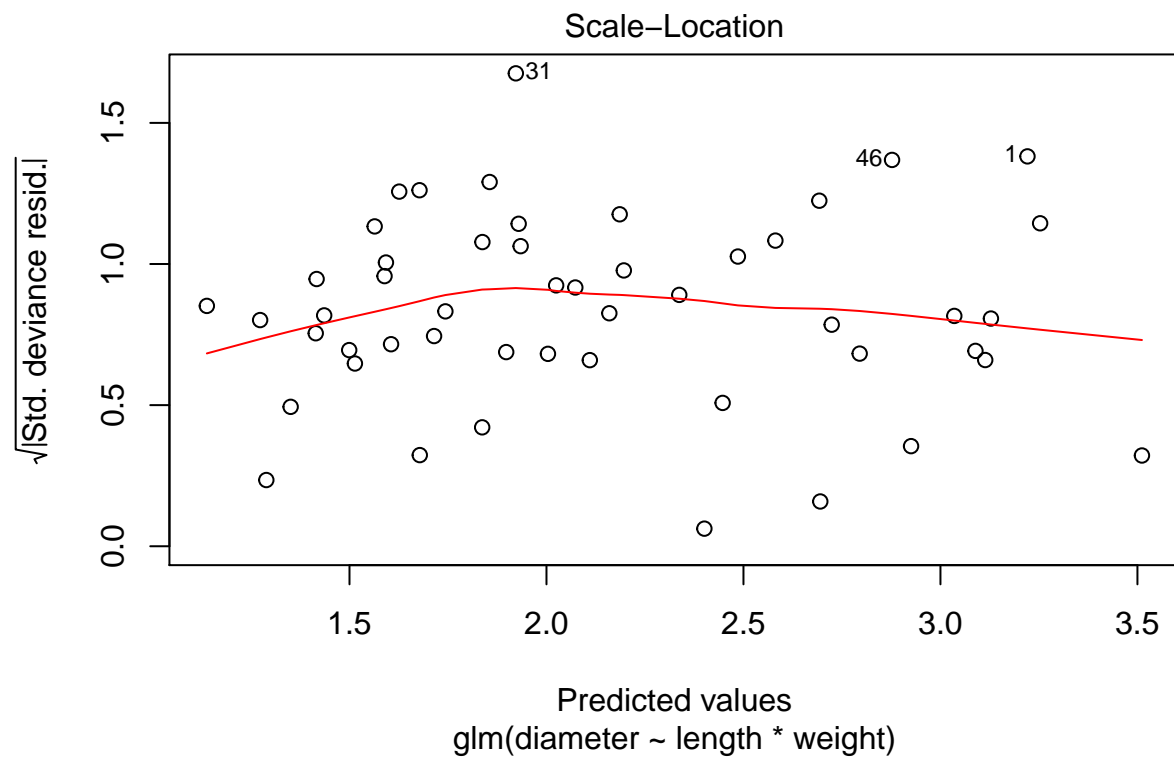
The top model is the glm with an interaction term between length and weight (glm_lw). Summary of model results.
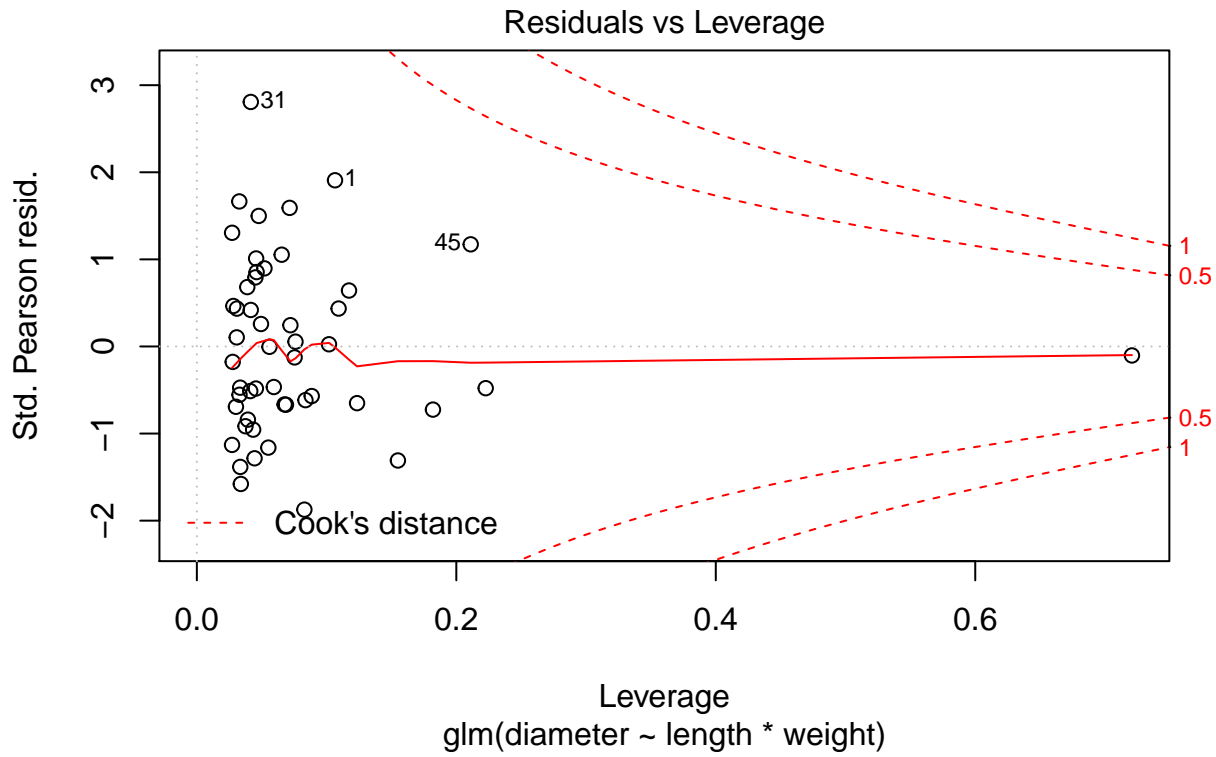
```
##
## Call:
## glm(formula = diameter ~ length * weight, data = hag)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.37718  -0.13558  -0.01846   0.13705   0.57776
##
## Coefficients:
##                  Estimate  Std. Error t value   Pr(>|t|)
## (Intercept)    1.04523103  0.28035763   3.728   0.000527 ***
## length        -0.00669661  0.01131453  -0.592   0.556843
## weight         0.02177107  0.00349228   6.234 0.000000129 ***
## length:weight -0.00024887  0.00004612  -5.396 0.000002304 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04419469)
##
##     Null deviance: 21.201  on 49  degrees of freedom
## Residual deviance:  2.033  on 46  degrees of freedom
## AIC: -8.2328
##
## Number of Fisher Scoring iterations: 2
```

Plots of model results

Residuals vs Fitted

Residuals

Predicted values
glm(diameter ~ length * weight)

Normal Q–Q

Theoretical Quantiles
glm(diameter ~ length * weight)

Scale–Location

√|Std. deviance resid.|

Predicted values
glm(diameter ~ length * weight)

Residuals vs Leverage

glm(diameter ~ length * weight)

Comparisons

Compare best fit model to the Harada estimate (a linear model) and a linear model using ADFG data. This was explored by examining model residuals by length and weight.

These show that the Harada parameter estimates do a poor job at lengths > 50, and are strongly biased when compared via weight. The best fit model does a good job of describing the output.

The coefficent estimates can now be used as predictors for future samples

```
length = seq(20, 60, by = 5)
length = 20
weight = 20
weight = seq(50, 450, by = 50)

coef(glm_lw)[[1]] + coef(glm_lw)[[2]] * length + coef(glm_lw)[[3]] * weight + coef(glm_lw)[[4]] * lengt
```

```
## [1] 1.750983 2.590667 3.430351 4.270035 5.109720 5.949404 6.789088 7.628772
## [9] 8.468456
```

```
summary(glm_lw)
```
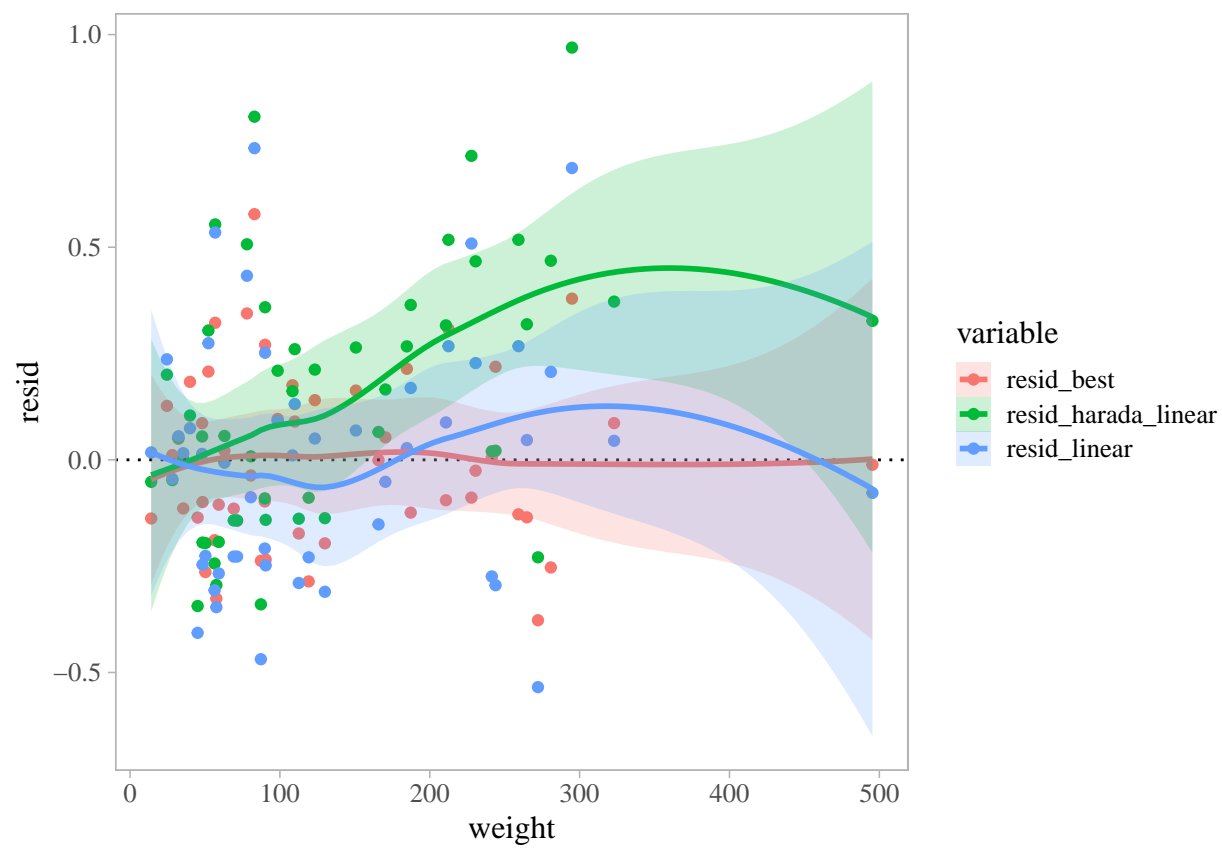
```
##
## Call:
```

9

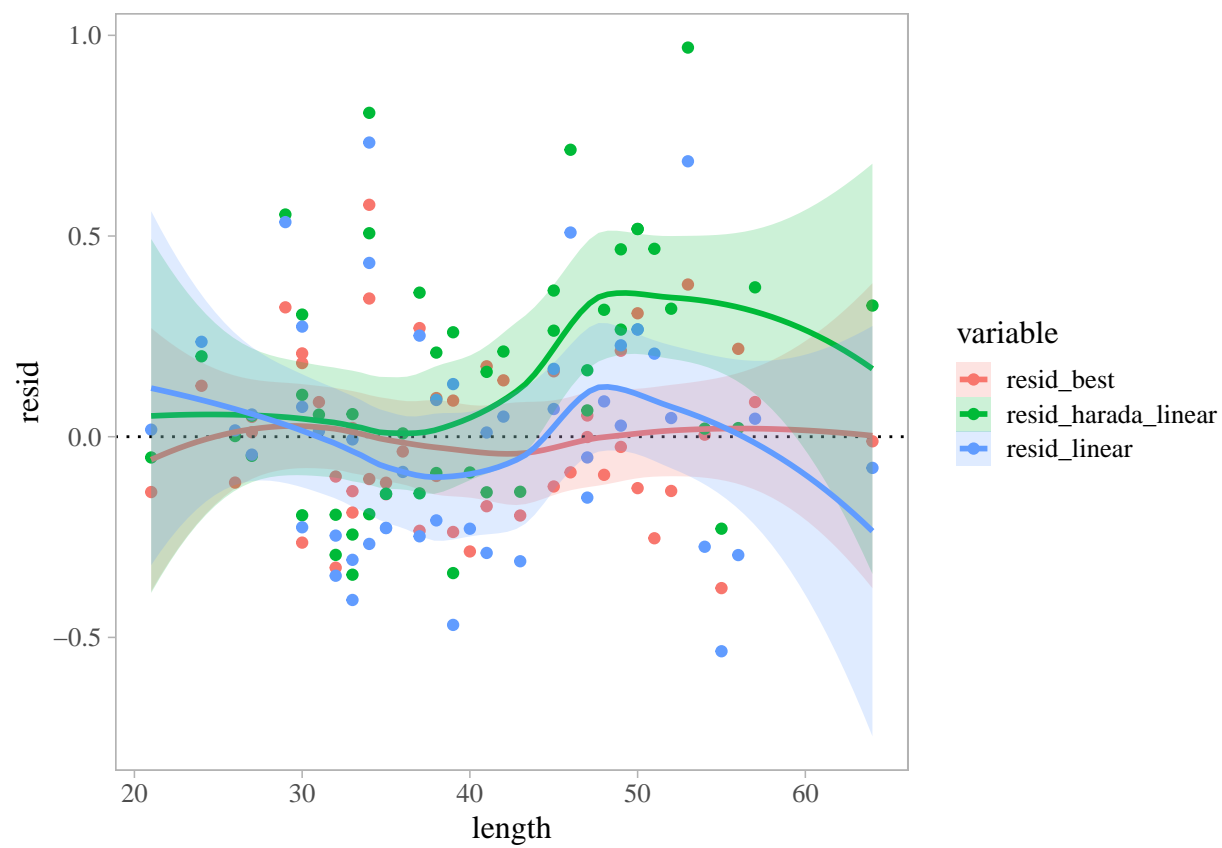Figure 4: Hagfish model residuals by weight.

Figure 5: Hagfish model residuals by length
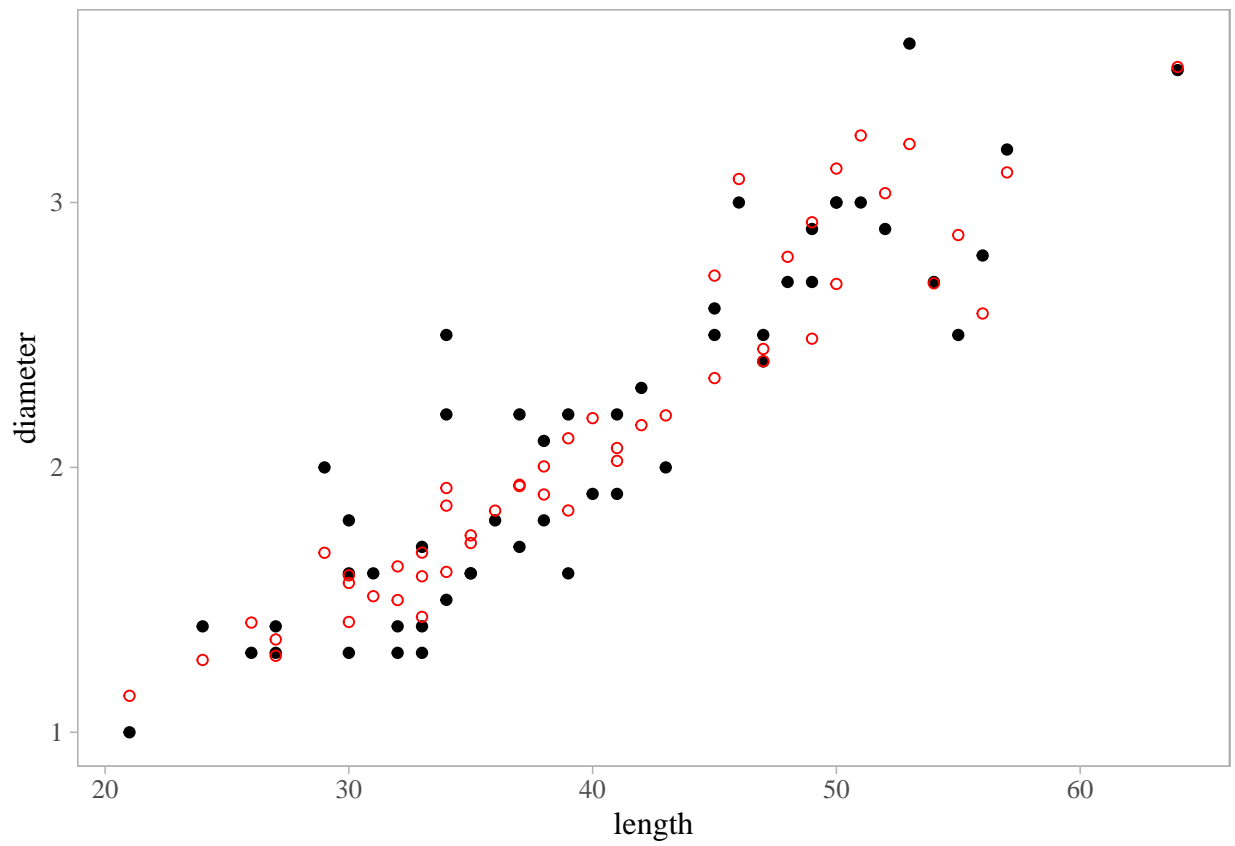
Figure 6: Observed and predicted hagfish diameters

```
## glm(formula = diameter ~ length * weight, data = hag)
##
## Deviance Residuals:
##      Min       1Q    Median       3Q      Max
## -0.37718  -0.13558  -0.01846   0.13705   0.57776
##
## Coefficients:
##                  Estimate  Std. Error t value    Pr(>|t|)
## (Intercept)    1.04523103  0.28035763   3.728    0.000527 ***
## length        -0.00669661  0.01131453  -0.592    0.556843
## weight         0.02177107  0.00349228   6.234 0.000000129 ***
## length:weight -0.00024887  0.00004612  -5.396 0.000002304 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04419469)
##
##     Null deviance: 21.201  on 49  degrees of freedom
## Residual deviance:  2.033  on 46  degrees of freedom
## AIC: -8.2328
##
## Number of Fisher Scoring iterations: 2
```

In short it appears that you should be able to use length/weight to get a reasonable estimate of diameter. This of course assumes that the sampling was distributed both in space and time - if not you will want to do some additional sampling to beef up this analysis.

Escape size

To determine escape size I worked up a function that predicts the L_{50} based upon 3 inputs:

- shrink = the amount a hagfish is estimated to reduce in size (e.g,. 0.15 is assuming a shrinkage of 15%) to fit through an opening
- escape_size = escape hole diameter (cm)
- prob = the probability e.g., (0.50 for 50%)

```
length_caught <- function(
  shrink, escape_size, prob) {

  hag %>%
    mutate(dia = diameter / (1 + shrink),
           catch = factor(ifelse(dia > escape_size, 1, 0))) -> df

    model = glm(catch ~ length, family = 'binomial', data = df)

    t = round((log(prob / (1 - prob)) - coef(model)[1]) / coef(model)[2])

    df %>%
      mutate(pred = predict(model, ., type ='response')) %>%
      ggplot(aes(length, pred)) + geom_point() +
      ggtitle(paste0('L', prob*100, ' = ', t, ' cm'))
}
```

```
length_caught(0.15, 1.9, 0.50)
```
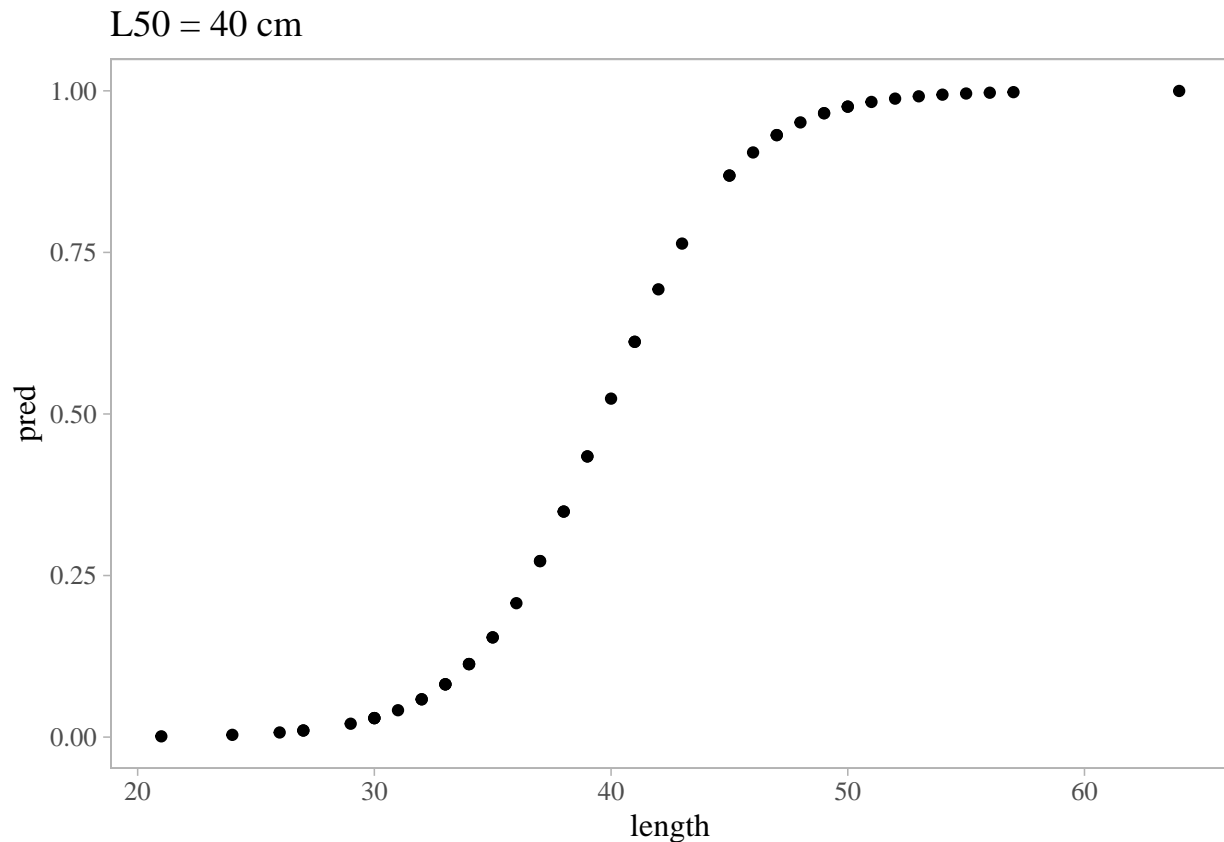

The length that 50% of the hagfish will be retained is reported at the top of the figure. The values can be adjusted in case you think that a hagfish can reduce in size by 20% etc.

Alternative method to get escape size

You could also use L_{50} (or some other reference point) from a maturity curve to determine an optimal escape hole size. Similar to above, this method is based on the estimated shrink, and the probability that you want to base the output on (defaults to 50%). In addition, you put in the L_{50} that is estimated externally from a maturity curve.

```
myL50 <- 45 # user input! waiting for the estimated L50 from from Andrew.

f_len_caught <- function(escape_size = 1.9,
                         shrink = 0.15,
                         prob = 0.5,
                         target_length = myL50) {
  hag %>% mutate(dia = diameter / (1 + shrink),
                 catch = factor(ifelse(dia > escape_size, 1, 0))) -> df
  model <- glm(catch ~ length, family = 'binomial', data = df)
  len <- round((log(prob / (1 - prob)) - coef(model)[1]) / coef(model)[2])
  out <- len - target_length # ultimately we want the difference between predicted len and target length
  return(out)
```

```
}

# Finds the escape ring size between the interval provided that gives us a 50% probability of releasing
fit <- uniroot(f = f_len_caught, interval = c(0.9,2.5))
optim_escape <- round(fit$root,1)
```

For example, given that the length at 50% maturity (L_{50}) is 45 cm, the optimal escape hole size that results in a 50% probability of retention for L_{50} is 2.2 cm.

Length predictions from the Harada model for comparison

As requested by Aaron, this function uses the paramter estimates for the logistic curve and linear regression from Harada et al. 2007 to output the length at which some proportion of fish (e.g. 50%) are retained given an escape hole size. The output is the length (cm) at which (prob, %) fish are retained by the gear given an escape hole size from the Harada model. Recall that the Harada model is based on relative girth, so length must be derived from the length/girth regression.

```
predict_harada <- function(
  # Probability from 0 to 1
  prob = 0.75,
  # Escape hole diameter (cm)
  escape_size = 1.9,
  # Logistic parameter estimates from Table 2 in Harada et al 2007
  logistic_alpha = -13.34,
  logistic_beta = 10.99,
  #length/girth linear regression estimates from Harada. Could use parameter
  #estimates from our best-fitting model instead.
  linreg_b0 = 4.213, # y-intercept
  linreg_b1 = 5.487 # slope
) {
# Relative girth (R) predicted by the model
R <- (log(prob/(1-prob))-logistic_alpha)/logistic_beta
# Girth (G) predicted by the model (adjusted by escape hole perimeter, which is diameter * pi)
G <- R * escape_size * pi
# Length predicted by the model using the length/girth linear regression estimates
L <- linreg_b0 + linreg_b1 * G
return(round(L,2))
}
```

For example, we can use this function to get the predicted lengths given some probability of retention and escape hole size:

```
prob_vec <- c(0.1, 0.25, 0.5, 0.75, 0.9)
escape_vec <- c(1.6, 1.9)

out <- matrix(nrow = length(prob_vec), ncol = length(escape_vec))

for(i in 1:length(prob_vec)){
  for(j in 1:length(escape_vec)){
out[i,j] <- predict_harada(prob = prob_vec[i], escape_size = escape_vec[j])
  }
}
```

```
out <- as.data.frame(out, row.names = FALSE)
names(out) <- paste0(escape_vec, " cm")
out %>%
  mutate(Probability = prob_vec) %>%
  select(Probability, `1.6 cm`, `1.9 cm`) -> out
```

Probability	1.6 cm	1.9 cm
0.10	32.18	37.42
0.25	34.93	40.69
0.50	37.69	43.97
0.75	40.45	47.24
0.90	43.21	50.52