

## Model Unveiling for Abalone Age

Casey Liu [csyliu@ucdavis.edu](mailto:csyliu@ucdavis.edu) 915216934

Zihan Xiao [zihxiao@ucdavis.edu](mailto:zihxiao@ucdavis.edu) 915532809

Yunwei Zhang 914652942

### Abstract:

In this project, we are fitting a regression model to predict abalone age using 8 different variables. Regression model was generated using training dataset and validated using validation dataset. AICp, BICp, Cp, SEEp, Pressp and MSPEv criterion was used for model selection and validation. X outlier and Y outliers was determined using leverage and Bonferroni method. Cook's distance method was used to identify influential outliers. As a result, the fitted regression model should simplify the process of abalone age prediction providing a more accurate and precise result.

### Introduction:

Abalone age can be approximated by counting number of rings by cutting shells through the cone. This method is time consuming, labor intensive, and error prone. However, we can use other measurements to predict age of abalone by fitting a regression model with respect to age. Abalone dataset has 4177 observations with 9 different variables (Sex: M, F, and I (infant), Length: Longest shell measurement (mm), Diameter: perpendicular to length (mm), Height: with meat in shell (mm), Whole weight: whole abalone (grams), Shucked weight: weight of meat (grams), Viscera weight: gut weight after bleeding (grams), Shell weight: after being dried (grams), Rings: +1.5 gives the age in years).

### Methods and Results:

Since we are interesting in age prediction, we calculated age by plus 1.5 to corresponded rings information in dataset. Age was chosen to be the response variable, the rest 8 variables (besides rings) are predictor variables. The data set was examined for missing values and improper column class, and neither of them was found. In other words, there is no need to replace missing value or convert column class for the original dataset, and it was ready for further analysis.

The distribution of response variable was first examined by drawing histogram (as shown in Plot 1a), which shows a right skewed tendency. Box-cox (as shown in Plot 1b) plot was used to decide which transformation we should use. Since lambda value was around 0, a log transformation to response variable is expected in order to achieve normal distribution for response variable. Histogram for transformed response variable was drawn (as shown in Plot 1c), and it shows a normal distribution with slight left skewness.

By analysing pairwise plot (as shown in Plot 1d), there is 8 quantitative variable (1 being response variable: Age, and 7 being predictor variables: Length, Height, Whole weight, Shucked weight, Viscera weight, Shell weight, Rings) and 1 qualitative (predictor: Sex) variable in the data set. Moreover, there is nonlinear relationship between some of the variables, thus second order should be considered for model fitting. More specifically, four weight related variables

(whole weight, shucked weight, viscera weight and shell weight) has nonlinear relationship with transformed response variables. Thus the fitted model should include second order of those variables. By examining correlation matrix for qualitative predictor variables (as shown in Plot 1e), all predictor variables are highly correlated, with each value larger than 0.7. Thus second order interaction should also be considered. Also notice that correlation coefficient between length and diameter are around 1, thus we assume they are perfectly correlated. In other words, by adding length variable in a model that already has diameter variable, length variable will not contribute anything towards data explanation, since two variables provide same information in data prediction. Thus only one of them is needed to be included in the model. We randomly chose length, rather than diameter, to be included in the model.

Box plot and pie chart for qualitative variable, sex variable, which has 3 levels (Female (F), Male (M) and Infant (I)), was drawn. From pie chart (as shown in Plot 2a), 3 levels have similar weight for population, since the area of pie chart for each level are almost the same. From box plot (as shown in Plot 2b), female and male have similar mean as well as same variance, since both boxes share similar size and similar mean indicator. However, notice that infant has slightly lower mean than the other two, but with similar box size. Thus a different mean between infant and adults (Male and Female) are expected, and factor effects should be taken into consideration when fitting models. Moreover, outliers were expected in those 3 levels, which may cause problems in model fitting.

The dataset was relatively larger than the amount of predictors, since there are 4177 observations and 7 predictors. It is reasonable for us to equally separate dataset into half for model validation in the future. The data was split into two groups, training group and validation group. When the data was split, there are 2089 observations in each data set and  $n/P$  will always be larger than 6 if  $P$  is smaller than 346. Thus a reliable model can be built based on training dataset. Model fitting, internal model validation will be conducted using training group, and external model validation will be conducted using validation group. By drawing boxplot for training data and validation data with respect to all variables side by side (as shown in Plot 3), we could see that there is similar distribution in these two sets, as they have similar mean and box size as well as outliers distribution for all variables. Thus by fitting models in training dataset, we should expect same prediction ability of chosen model when fitting validation dataset.

We are using AIC forward stepwise to fit training data with all first order effects (as shown in Plot 4a), and the chosen model was denoted as first model. Function of the first model is (rings+1.5=age):

$$\begin{aligned} \log(\text{Rings} + 1.5) = & 1.72618 + 0.74659 \times \text{shell weight} - 1.46747 \times \text{shucked weight} + 1.17087 \times \text{length} \\ & - 0.09120 \times \text{sex}(\text{Infant}) + 0.01104 \times \text{sex}(\text{Male}) + 0.80398 \times \text{height} + 0.51090 \times \text{whole weight} \\ & - 0.54916 \times \text{viscera weight} + \varepsilon \end{aligned}$$

The first model 8 regression coefficient (with 2 of them being 2 different levels of one factor variable) of 7 predictors was included. Notice that in those variables, sexM is insignificant

at  $\alpha$  level of 0.05. This is caused by the similar distribution between female and male distribution with respect to response variable. However, we decide to keep both regression coefficient. Since T test is only testing marginal effect of whether an predictor variable is significant given all other predictor variables being included in the model. This test of marginal effect of predictor variables in first model alone does not equivalent to testing whether there is a regression relation between response variable and sex M. The reduced model for each individual test contains all other predictor variables and thus may lead to non-significant results due to multicollinearity. Moreover, by examining F test for first order model using Anova tabling (as shown in Plot 4b), all regression coefficient are significant at  $\alpha$  level of 0.05. Here, the reduced model for testing regression relation contains no predictor variables. Thus all regression coefficient in the first order is important and are needed to be included for further analysis.

As for the first model, the residual vs fitted value for first model (as shown in Plot 4c) shows equal spread of residuals with different fitted value, thus linearity assumption of linear regression was hold. However the line on residual plot is not horizontal but is around zero, thus constant variance assumption of linear regression might not hold. Although the distribution is a little heavy tailed compared to normal distribution based on Q-Q plot (as shown in Plot 4d), Normality of the errors assumption is hold since the distribution in Q-Q plot is symmetric and the majority point are nearly on a straight line.

Based on the result of first order model fitting, we kept all 8 regression coefficient. Second order effects (for four weight related variables) and interaction effects between variable was also taken into consideration. By fitting the model with all first order, 2 way interaction effects of 7 predictors and four second order of 4 predictors, there is 49 coefficient, which is smaller than number of observations, in the fitted model. Then, AIC forward stepwise procedure was performed on this full model(as shown in Plot 5a). The result from this fitting process was denoted as the second model.

The function of the second model:

$$\begin{aligned} \log(Rings + 1.5) = & 1.426616 + 4.185854 \times shell\ weight + 3.408829 \times shucked\ weight^2 \\ & - 1.274031 \times shell\ weight^2 - 1.621422 \times shucked\ weight - 0.273622 \times sex(Infant) \\ & - 0.117077 \times sex(Male) - 0.544771 \times whole\ weight + 1.917785 \times length + 6.885656 \times height \\ & + 0.402232 \times viscera\ weight + 0.330107 \times shucked\ weight \times sexI - 0.003283 \times shucked\ weight \times sex(Male) \\ & - 2.307068 \times shucked\ weight \times whole\ weigh - 7.323386 \times shell\ weight \times length - 0.457542 \times \\ & shell\_weight \times whole\ weight + 0.878959 \times sexI \times height + 0.800871 \times sexM \times height - 12.451029 \\ & \times shucked\ weight \times height + 3.351807 \times shell\ weight \times height - 15.836451 \times length \times height + 2.394141 \\ & \times whole\ weight \times length + 3.854407 \times shell\ weight \times shucked\ weight + 6.960776 \times whole\ weight \times height + \varepsilon \end{aligned}$$

However, since the second model has 23 regression coefficient, it needs large amount of information, and might be potentially overfitting and hard to analysis. Recall that BIC criterion

is puts more penalty on model complexity, thus favors smaller model than AIC. Under this circumstances, BIC forward stepwise procedure was performed on the full model in order to get a smaller regression model, and the selected model was denoted as the third model.

Function of the third model:

$$\begin{aligned} \log(\text{Rings} + 1.5) = & 1.72041 + 3.63401 \times \text{shell weight} + 2.74745 \times \text{shucked weight}^2 - 2.04587 \times \text{shell weight}^2 \\ & - 2.79236 \times \text{shucked weight} - 0.19090 \times \text{sex}(\text{Infant}) + 0.02607 \times \text{sex}(\text{Male}) - 0.85304 \times \text{whole weight} \\ & + 1.19142 \times \text{length} + 0.36942 \times \text{height} + 0.45644 \times \text{shucked weight} \times \text{sex}(\text{Infant}) + 0.07468 \times \text{shucked weight} \\ & \times \text{sex}(\text{Male}) - 1.39382 \times \text{shucked weight} \times \text{whole weight} - 4.72450 \times \text{shell weight} \times \text{length} \\ & + 1.34892 \times \text{shell weight} \times \text{whole weight} + \varepsilon \end{aligned}$$

The third model only contains 14 regression coefficients(as shown in Plot 5b). Same as the first model, there is some regression coefficient being insignificant at  $\alpha$  level of 0.05 in both second and third model. However, since T test is only testing reduced model with all rest available predictor variables versus full model, it cannot tell whether there is a regression relation between response variable and those predictors. By examining F test for the second model using Anova tabling (as shown in Plot 5c for second model and Plot 5d for third model), all regression coefficient are significant at  $\alpha$  level of 0.05. Here, the reduced model for testing regression relation contains no predictor variables. Thus all regression coefficient in the second order is important based on AIC or BIC analysis and are needed to be included for further analysis.

As for the second and third model, the residual vs fitted value (as shown in Plot 5e for second model and Plot 5f for third model ) show equal spread of residuals with different fitted value, and the fitted line are more horizontal and closer to zero compare to the first model. Thus linearity and constant variance assumption of linear regression was hold. Q-Q plot shows a less heavy tailed distribution and more points land on straight line compared to the first model based. Thus, normality of the errors assumption is hold.

The final model need to be decided among the first, second and third model. Cp, SSEp, adjusted R square, MSEp, MSPEv value was calculated for each of these three model (as shown in Plot 6a). Since model 2 and model 3 has similar value in both of those SSE, MSE, MSPEv, AIC and BIC value, both of them explains the data in the same level. Moreover, model 2 has a Cp value closer to corresponding p value. From the table we know that model 1 with least variables have the largest cp, which means this model has large amounts of bias. Thus we should not choose this one. Based on adjusted R square, AIC and BIC criterion, model 3 is favored since it has the lowest BIC value. Based on adjusted R square, model 2 is favored since it has the highest adjusted R square. Based on Cp, model 2 is favored since even with the most variables, it has the smallest cp value, which means it has the least model bias and least msee by using bias-variance trade off. Thus model 2 explain training dataset more thoroughly with less degrees of freedom. However, since adjusted R and AIC BIC value did not change very much between model 2 and model 3 while these two model have a large difference in cp, we should take cp as the most important and crucial factor.

Based on above analysis, the second model was chosen as the final model. Then we find that by applying same variables into validation data and using linear regression method to fit the data(as shown in Plot 6b), the new fitted model has some regression coefficients with opposite signs to the signs of corresponding estimators in model generated by using training data. This phenomenon shows that model 2 is kind of overfitted. However, recall that MSPEv value for the second model is not much larger than  $SSE/n$ , it's not severe overfitting.

Since the final model has been validated and chosen, the entire data set was used to re-fit the final model(as shown in plot 7a). The residual plot and Q-Q plot was drawn(as shown in 7b). Both of the three normal assumptions was hold by examining these two plots. However, on Q-Q plot we could notice some outliers was on both tail, thus outlier analysis should be taken into consideration.

X outliers was identified using leverage method (as shown in Plot 8a), and the residual versus leverage plot was drawn (as shown in 8b). There is a lot X residuals, thus influential cases might be identified among them. Y outliers was identified using studentized deleted residuals and Bonferroni outlier test procedure at  $\alpha = 0.1$  (as shown in 8c). Since the largest studentized deleted residuals is smaller than the Bonferroni threshold value. Thus, there is no Y outliers and large amount of X outliers based on those criterion. Influence index plot using Cook's distance was drawn (as shown in 8d), and on influence index plot we could see that 3 cases has high cook's distance value compared with the rest of data. Then we calculate the largest cook's distance and we find that 2052th case is the largest influential case in full dataset with 0.99 probability higher than threshold.

By removing influential case from original dataset, the dataset was fitted the final model(as shown in 9a). The residual plot and Q-Q plot was drawn(as shown in 9b). Both linearity, normality, constant variance assumption are hold. Moreover, there is no more influential cases (as shown in 9c).

## Conclusions and Discussion:

The final model is chosen, and the function is:

$$\begin{aligned} \log(Rings + 1.5) = & 1.44681 + 3.55769 \times shell\ weight + 3.24733 \times shucked\ weight^2 \\ & - 0.44735 \times shell\ weight^2 - 1.92368 \times shucked\ weight - 0.26375 \times sex(Infant) \\ & - 0.10759 \times sex(Male) - 0.07583 \times whole\ weight + 1.80799 \times length + 6.32274 \times height \\ & - 0.53894 \times viscera\ weight + 0.32270 \times shucked\ weight \times sex(Infant) - 0.02531 \times shucked\ weight \\ & \times sex(Male) - 1.86049 \times shucked\ weight \times whole\ weight - 5.07462 \times shell\ weight \times length - 0.43845 \times \\ & shell\ weight \times whole\ weight + 0.87329 \times sexI \times height + 0.76494 \times sex(Male) \times height - 9.18711 \\ & \times shucked\ weight \times height + 0.43878 \times shell\ weight \times height - 14.29705 \times length \times height + 1.64716 \\ & \times whole\ weight \times length + 2.11963 \times shell\ weight \times shucked\ weight + 5.75455 \times whole\ weight \times height + \epsilon \end{aligned}$$

Notice viscera weight only has one term in the final function, while as the other predictor has at least two terms. It is reasonable since the difference age of abalone should not shows a difference of weight of viscera but the condition of viscera. In other words, when Abalone grows older, we should expect the functioning ability change in viscera, by analysing weight itself, does not contribute much towards age prediction. Moreover, diameter was deleted at the beginning since it is almost perfect correlated with length, we only need to keep one of them in model. This is reasonable since length and diameter are both measuring size of abalone. Besides viscera weight and diameter, all the other predictor contributes a lot into data explanation of the final model.

Notice that the final model has both first model, second model and 2-way interaction terms, which corresponds to our first analysis based on pairwise interaction plot (Plot 1d). Also different level of sex contributes to different significant regression coefficient, thus factor effects was included in the final model.

Although we want to fit the data into a smaller model either by BIC stepwise procedure(the third model) or fitting model based on first order(the first model), the second model has less variance with little bias though it has the most regression coefficients. However, the second model is overfitting, since the sign of regression coefficient was changed when fitting the model to validation dataset. But, the second model still has a MSPEv value close to  $SSE/n$  and  $Pressp/n$ , which means it did not severely overfitting. Thus it is reasonable for us to say the second model is the best model among those three.

However, since we used BIC criterion to the second full model (with second order of all weight related predictors along with first order and 2-way interaction of all predictors), we should also conduct BIC to the first full model(with first order of all predictors) to generate a fourth model. By doing this, it will provide a fair competition between first order model and second order 2 way interaction model. Moreover, since we jumped from first order right onto both second order and interaction model, it is reasonable for us to see a drastic model fitting as well as model complexity change between them. In this case, we should also fit two model only consider first and second order by using AIC and BIC criterion, and fit another two model only consider first and 2-way interaction by using AIC and BIC criterion. By doing this, we should have more option while model selection and validation.

Appendix:

1.)Model Function :

Model 1 :

$$\begin{aligned} \log(Rings + 1.5) = & 1.72618 + 0.74659 \times shell\ weight - 1.46747 \times shucked\ weight + 1.17087 \times length \\ & - 0.09120 \times sex(Infant) + 0.01104 \times sex(Male) + 0.80398 \times height + 0.51090 \times whole\ weight \\ & - 0.54916 \times viscera\ weight + \varepsilon \end{aligned}$$

Model 2:

$$\begin{aligned} \log(Rings + 1.5) = & 1.426616 + 4.185854 \times shell\ weight + 3.408829 \times shucked\ weight^2 \\ & - 1.274031 \times shell\ weight^2 - 1.621422 \times shucked\ weight - 0.273622 \times sex(Infant) \\ & - 0.117077 \times sex(Male) - 0.544771 \times whole\ weight + 1.917785 \times length + 6.885656 \times height \\ & + 0.402232 \times viscera\ weight + 0.330107 \times shucked\ weight \times sex(Infant) - 0.003283 \times shucked\ weight \\ & \times sex(Male) - 2.307068 \times shucked\ weight \times whole\ weight - 7.323386 \times shell\ weight \times length - 0.457542 \times \\ & shell\ weight \times whole\ weight + 0.878959 \times sexI \times height + 0.800871 \times sex(Male) \times height - 12.451029 \\ & \times shucked\ weight \times height + 3.351807 \times shell\ weight \times height - 15.836451 \times length \times height + 2.394141 \\ & \times whole\ weight \times length + 3.854407 \times shell\ weight \times shucked\ weight + 6.960776 \times whole\ weight \times height + \varepsilon \end{aligned}$$

Model 3:

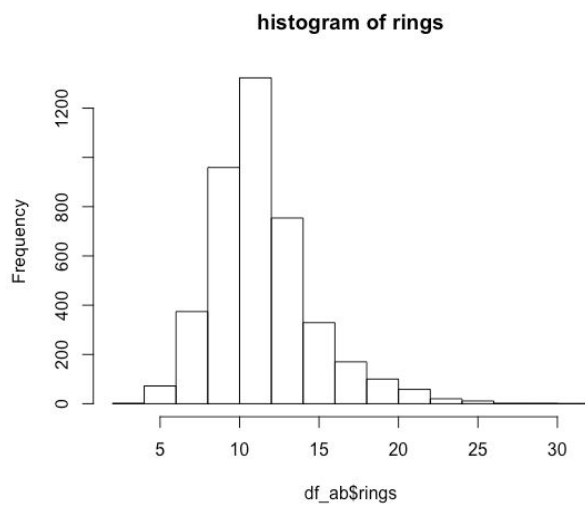
$$\begin{aligned} \log(Rings + 1.5) = & 1.72041 + 3.63401 \times shell\ weight + 2.74745 \times shucked\ weight^2 - 2.04587 \times shell\ weight^2 \\ & - 2.79236 \times shucked\ weight - 0.19090 \times sex(Infant) + 0.02607 \times sex(Male) - 0.85304 \times whole\ weight \\ & + 1.19142 \times length + 0.36942 \times height + 0.45644 \times shucked\ weight \times sexI + 0.07468 \times shucked\ weight \\ & \times sex(Male) - 1.39382 \times shucked\ weight \times whole\ weight - 4.72450 \times shell\ weight \times length \\ & + 1.34892 \times shell\ weight \times whole\ weight + \varepsilon \end{aligned}$$

Final Model:

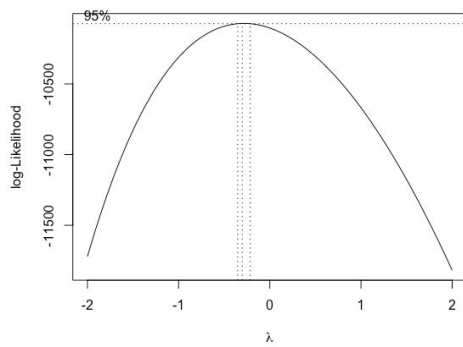
$$\begin{aligned} \log(Rings + 1.5) = & 1.44681 + 3.55769 \times shell\ weight + 3.24733 \times shucked\ weight^2 \\ & - 0.44735 \times shell\ weight^2 - 1.92368 \times shucked\ weight - 0.26375 \times sex(Infant) \\ & - 0.10759 \times sex(Male) - 0.07583 \times whole\ weight + 1.80799 \times length + 6.32274 \times height \\ & - 0.53894 \times viscera\ weight + 0.32270 \times shucked\ weight \times sex(Infant) - 0.02531 \times shucked\ weight \\ & \times sex(Male) - 1.86049 \times shucked\ weight \times whole\ weight - 5.07462 \times shell\ weight \times length - 0.43845 \times \\ & shell\ weight \times whole\ weight + 0.87329 \times sexI \times height + 0.76494 \times sex(Male) \times height - 9.18711 \\ & \times shucked\ weight \times height + 0.43878 \times shell\ weight \times height - 14.29705 \times length \times height + 1.64716 \\ & \times whole\ weight \times length + 2.11963 \times shell\ weight \times shucked\ weight + 5.75455 \times whole\ weight \times height + \varepsilon \end{aligned}$$

## 2.)Plots

Plot 1a.



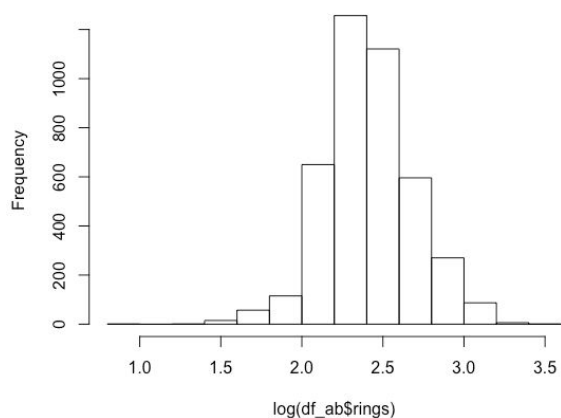
Plot 1b.)



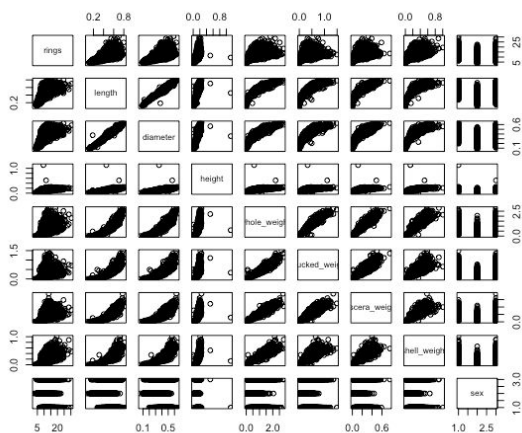
Plot 1c.)



histogram of response variable with log transformation



Plot 1d.)



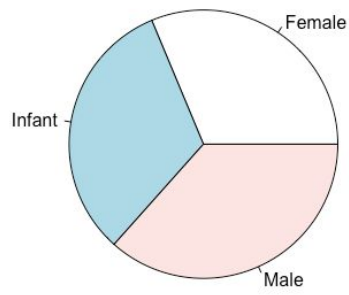
Plot 1e.)

```
> correlation_matrix
```

	length	diameter	height	whole_weight	shucked_weight	viscera_weight	shell_weight
length	1.0000000	0.9868116	0.8275536	0.9252612	0.8979137	0.9030177	0.8977056
diameter	0.9868116	1.0000000	0.8336837	0.9254521	0.8931625	0.8997244	0.9053298
height	0.8275536	0.8336837	1.0000000	0.8192208	0.7749723	0.7983193	0.8173380
whole_weight	0.9252612	0.9254521	0.8192208	1.0000000	0.9694055	0.9663751	0.9553554
shucked_weight	0.8979137	0.8931625	0.7749723	0.9694055	1.0000000	0.9319613	0.8826171
viscera_weight	0.9030177	0.8997244	0.7983193	0.9663751	0.9319613	1.0000000	0.9076563
shell_weight	0.8977056	0.9053298	0.8173380	0.9553554	0.8826171	0.9076563	1.0000000

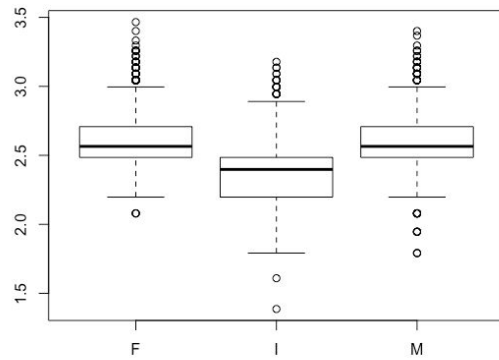
Plot 2a.)

pie plot for sex



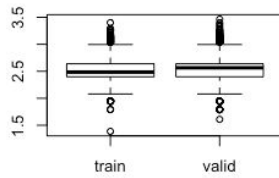
Plot 2b.)

boxplot of age by sex

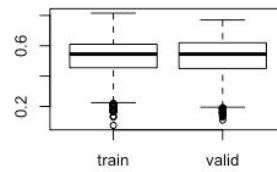


Plot 3.)

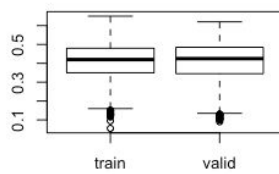
boxplot of log\_age by data type



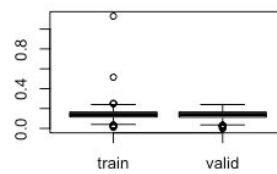
boxplot of length by data type



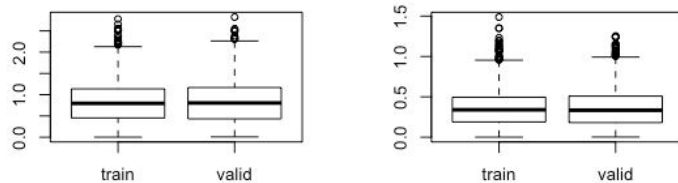
boxplot of diameter by data type



boxplot of height by data type



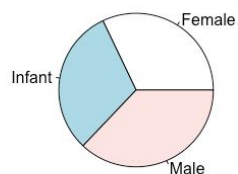
boxplot of whole\_weight by data type boxplot of shucked\_weight by data type



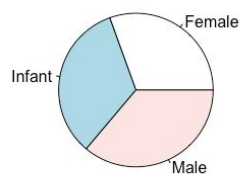
boxplot of viscera\_weight by data type boxplot of shell\_weight by data type



pie plot for sex in training data



pie plot for sex in validation data



## Plot 4a)

Step: AIC=-7784.46  
 $\log\_age \sim \text{shell\_weight} + \text{shucked\_weight} + \text{length} + \text{sex} + \text{height} + \text{whole\_weight} + \text{viscera\_weight}$

	Df	Sum of Sq	RSS	AIC
<none>			49.870	-7784.5
- viscera_weight	1	0.3335	50.203	-7772.5
- height	1	0.7643	50.634	-7754.7
- shell_weight	1	0.8492	50.719	-7751.2
- whole_weight	1	0.9822	50.852	-7745.7
- sex	2	2.2429	52.113	-7696.6
- length	1	3.4429	53.313	-7647.0
- shucked_weight	1	6.3045	56.174	-7537.8

Call:  
 $\text{lm}(\text{formula} = \log\_age \sim \text{shell\_weight} + \text{shucked\_weight} + \text{length} + \text{sex} + \text{height} + \text{whole\_weight} + \text{viscera\_weight}, \text{data} = \text{train})$

Coefficients:

	shell_weight	shucked_weight	length	sexI	sexM
(Intercept)	1.963380	0.678683	-1.311051	0.961040	-0.079075
height	0.698724	0.468222	-0.490188		0.009785

```
> summary(model_fit1)
```

Call:  
lm(formula = log\_age ~ shell\_weight + shucked\_weight + length + sex + height + whole\_weight + viscera\_weight, data = train)

Residuals:

	Min	1Q	Median	3Q	Max
	-0.66891	-0.09847	-0.01792	0.07726	0.61617

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.963380	0.029497	66.563	< 2e-16 ***
shell_weight	0.678683	0.114036	5.951	3.11e-09 ***
shucked_weight	-1.311051	0.080850	-16.216	< 2e-16 ***
length	0.961040	0.080199	11.983	< 2e-16 ***
sexI	-0.079075	0.010038	-7.878	5.33e-15 ***
sexM	0.009785	0.008222	1.190	0.234183
height	0.698724	0.123758	5.646	1.87e-08 ***
whole_weight	0.468222	0.073156	6.400	1.91e-10 ***
viscera_weight	-0.490188	0.131423	-3.730	0.000197 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1548 on 2080 degrees of freedom  
Multiple R-squared: 0.5738, Adjusted R-squared: 0.5722  
F-statistic: 350.1 on 8 and 2080 DF, p-value: < 2.2e-16

Plot 4b.)

```
> anova(model_fit1)
```

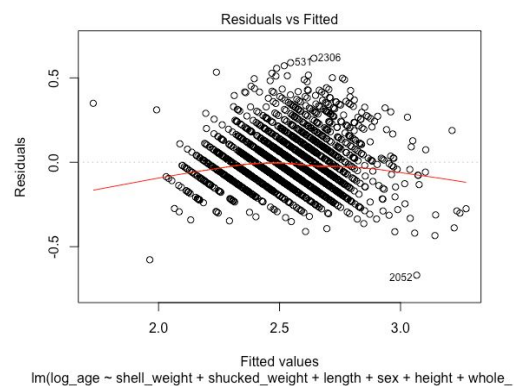
Analysis of Variance Table

Response: log\_age

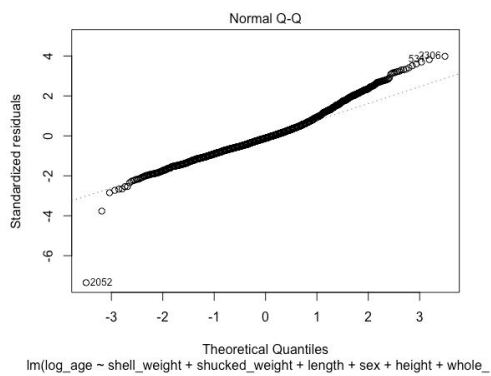
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
shell_weight	1	49.802	49.802	2077.167	< 2.2e-16 ***
shucked_weight	1	7.185	7.185	299.685	< 2.2e-16 ***
length	1	5.834	5.834	243.319	< 2.2e-16 ***
sex	2	2.587	1.294	53.958	< 2.2e-16 ***
height	1	0.761	0.761	31.756	1.986e-08 ***
whole_weight	1	0.650	0.650	27.110	2.111e-07 ***
viscera_weight	1	0.334	0.334	13.912	0.0001967 ***
Residuals	2080	49.870	0.024		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Plot 4c.)



Plot 4d.)



Plot 5a.)

Step: AIC=-8093.98

log\_age ~ shell\_weight + I(shucked\_weight \* shucked\_weight) +  
I(shell\_weight \* shell\_weight) + shucked\_weight + sex + whole\_weight +  
height + viscera\_weight + length + shucked\_weight:sex + shucked\_weight:whole\_weight +  
sex:height + shucked\_weight:height + whole\_weight:height +  
height:length + shell\_weight:shucked\_weight + shucked\_weight:length +  
shell\_weight:length + whole\_weight:length

	Df	Sum of Sq	RSS	AIC
<none>			42.430	-8094.0
+ I(viscera_weight * viscera_weight)	1	0.028569	42.401	-8093.4
+ shucked_weight:viscera_weight	1	0.020174	42.410	-8093.0
+ length:sex	2	0.057907	42.372	-8092.8
+ height:shell_weight	1	0.015886	42.414	-8092.8
+ whole_weight:shell_weight	1	0.014773	42.415	-8092.7
+ length:viscera_weight	1	0.012252	42.417	-8092.6
+ whole_weight:viscera_weight	1	0.012195	42.417	-8092.6
+ height:viscera_weight	1	0.012091	42.418	-8092.6
+ shell_weight:sex	2	0.047239	42.382	-8092.3
+ I(whole_weight * whole_weight)	1	0.000566	42.429	-8092.0
+ shell_weight:viscera_weight	1	0.000084	42.430	-8092.0
+ whole_weight:sex	2	0.027756	42.402	-8091.3
+ viscera_weight:sex	2	0.019743	42.410	-8091.0

Plot 5b.)

```
> summary(model_fit2)

Call:
lm(formula = log_age ~ shell_weight + I(shucked_weight * shucked_weight) +
    I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +
    length + height + viscera_weight + shucked_weight:sex + shucked_weight:whole_weight +
    shell_weight:length + shell_weight:shell_weight + sex:height +
    shucked_weight:height + shell_weight:height + length:height +
    whole_weight:length + shell_weight:shucked_weight + whole_weight:height,
    data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-0.53408 -0.10409 -0.01158  0.08679  0.65231

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.426616    0.102917  13.862 < 2e-16 ***
shell_weight    4.185854    1.219843   3.431 0.000612 ***
I(shucked_weight * shucked_weight) 3.408829    0.463337   7.357 2.70e-13 ***
I(shell_weight * shell_weight)    -1.274031    1.227595  -1.038 0.299472
shucked_weight  -1.621422    0.432138  -3.752 0.000180 ***
sexI            -0.273622    0.047611  -5.747 1.04e-08 ***
sexM            -0.117077    0.043298  -2.704 0.006908 **
whole_weight    -0.544771    0.426017  -1.279 0.201128
length          1.917785    0.289448   6.626 4.40e-11 ***
height          6.885656    1.395066   4.936 8.63e-07 ***
viscera_weight  -0.402232    0.141322  -2.846 0.004468 **
shucked_weight:sexI  0.330107    0.113544   2.907 0.003684 **
shucked_weight:sexM -0.003283    0.059666  -0.055 0.956127
shucked_weight:whole_weight -2.307068    0.438396  -5.263 1.57e-07 ***
shell_weight:length -7.323386    2.583473  -2.835 0.004631 **
shell_weight:whole_weight -0.457542    0.823140  -0.556 0.578374
sexI:height      0.878959    0.498882   1.762 0.078242 .
sexM:height      0.800871    0.371259   2.157 0.031107 *
shucked_weight:height -12.451029    4.041368  -3.081 0.002091 **
shell_weight:height  3.351807    6.024556   0.556 0.578027
length:height    -15.836451    3.753701  -4.219 2.50e-05 ***
whole_weight:length  2.304141    0.607652   3.432 0.000612 ***
shell_weight:shucked_weight  3.854407    1.228388   3.138 0.001726 **
whole_weight:height  6.960776    3.231311   2.154 0.031343 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1615 on 2065 degrees of freedom
Multiple R-squared:  0.6448,    Adjusted R-squared:  0.6408
```

## Plot 5c.)

```
> anova(model_fit2)

Analysis of Variance Table

Response: log_age

            Df Sum Sq Mean Sq  F value    Pr(>F)
shell_weight    1 65.011  65.011 2492.0836 < 2.2e-16 ***
I(shucked_weight * shucked_weight) 1 11.900  11.900  456.1626 < 2.2e-16 ***
I(shell_weight * shell_weight)    1  4.030   4.030  154.4652 < 2.2e-16 ***
shucked_weight    1  4.745   4.745  181.8853 < 2.2e-16 ***
sex              2  1.888   0.944   36.1820 3.593e-16 ***
whole_weight     1  1.279   1.279  49.0262 3.403e-12 ***
length          1  1.201   1.201  46.0340 1.513e-11 ***
height          1  0.644   0.644  24.7034 7.237e-07 ***
viscera_weight  1  0.552   0.552  21.1599 4.481e-06 ***
shucked_weight:sex  2  2.204   1.102  42.2504 < 2.2e-16 ***
shucked_weight:whole_weight  1  0.985   0.985  37.7419 9.661e-10 ***
shell_weight:length  1  1.479   1.479  56.7001 7.540e-14 ***
shell_weight:whole_weight  1  0.209   0.209   7.9975 0.0047294 **
sex:height       2  0.327   0.163   6.2601 0.0019475 **
shucked_weight:height  1  0.331   0.331  12.6922 0.0003755 ***
shell_weight:height  1  0.226   0.226   8.6764 0.0032596 **
length:height    1  0.295   0.295  11.2996 0.0007894 ***
whole_weight:length  1  0.215   0.215   8.2531 0.0041098 **
shell_weight:shucked_weight  1  0.137   0.137   5.2687 0.0218117 *
whole_weight:height  1  0.121   0.121   4.6404 0.0313427 *
Residuals      2065 53.870   0.026
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Plot 5d.)



```

Step: AIC=-7955.08
log_age ~ shell_weight + I(shucked_weight * shucked_weight) +
  I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +
  height + viscera_weight + shucked_weight:sex + shucked_weight:whole_weight +
  shucked_weight:height + sex:height + whole_weight:height

Df Sum of Sq    RSS      AIC
<none>                 43.561 -7955.1
+ length               1  0.074963 43.486 -7951.0
+ height:viscera_weight 1  0.017808 43.543 -7948.3
+ I(viscera_weight * viscera_weight) 1  0.016693 43.544 -7948.2
+ I(whole_weight * whole_weight) 1  0.016299 43.544 -7948.2
+ height:shell_weight 1  0.012394 43.548 -7948.0
+ shucked_weight:viscera_weight 1  0.008156 43.552 -7947.8
+ shell_weight:shucked_weight 1  0.006805 43.554 -7947.8
+ whole_weight:viscera_weight 1  0.006805 43.555 -7947.7
+ shell_weight:viscera_weight 1  0.000798 43.560 -7947.5
+ whole_weight:shell_weight 1  0.000501 43.560 -7947.5
+ shell_weight:sex      2  0.091968 43.469 -7944.2
+ whole_weight:sex      2  0.063085 43.497 -7942.8
+ viscera_weight:sex    2  0.005688 43.555 -7940.1

Call:
lm(formula = log_age ~ shell_weight + I(shucked_weight * shucked_weight) +
  I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +
  height + viscera_weight + shucked_weight:sex + shucked_weight:whole_weight +
  shucked_weight:height + sex:height + whole_weight:height,
  data = train)

Coefficients:
            (Intercept)                shell_weight  I(shucked_weight * shucked_weight)
                2.25008                      1.36994                      2.49192
I(shell_weight * shell_weight)          shucked_weight                sexI
   -0.97492                    -1.44294                   -0.34826
sexM                    whole_weight                height
   -0.16149                    0.65046                   1.34675
viscera_weight          shucked_weight:sexI          shucked_weight:sexM
   -0.39387                    0.20945                   -0.03681
shucked_weight:whole_weight  shucked_weight:height  sexI:height
   -0.95726                    -8.55614                   1.77637
sexM:height                whole_weight:height
    1.18966                    2.72114

> summary(model_fit3)

Call:
lm(formula = log_age ~ shell_weight + I(shucked_weight * shucked_weight) +
  I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +
  length + height + shucked_weight:sex + shucked_weight:whole_weight +
  shell_weight:length + shell_weight:whole_weight, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-0.49546 -0.09170 -0.01310  0.07616  0.57841

Coefficients:
            (Intercept)                Estimate Std. Error t value Pr(>|t|)
shell_weight              3.09698          0.31665    9.780 < 2e-16 ***
I(shucked_weight * shucked_weight) 2.45497          0.37745    6.504 9.77e-11 ***
I(shell_weight * shell_weight)    -1.76994          0.72168   -2.453  0.01427 *
shucked_weight              -2.48762          0.16055   -15.495 < 2e-16 ***
sexI                    -0.16635          0.02096   -7.937 3.36e-15 ***
sexM                    -0.02387          0.01785   -1.337  0.18125
whole_weight              0.79067          0.11490    6.881 7.83e-12 ***
length                    0.93923          0.13218    7.106 1.64e-12 ***
height                    0.32259          0.11852    2.722  0.00655 **
shucked_weight:sexI        0.40015          0.06512    6.145 9.55e-10 ***
shucked_weight:sexM        0.06846          0.03712    1.845  0.06523 .
shucked_weight:whole_weight -1.25315          0.28767   -4.356 1.39e-05 ***
shell_weight:length        -3.91712          0.49903   -7.849 6.64e-15 ***
shell_weight:whole_weight   1.14786          0.42337    2.711  0.00676 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1452 on 2074 degrees of freedom
Multiple R-squared:  0.6263,    Adjusted R-squared:  0.6237
F-statistic: 248.2 on 14 and 2074 DF,  p-value: < 2.2e-16

```

```

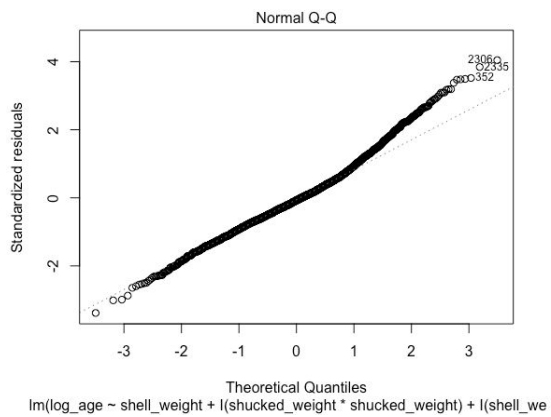
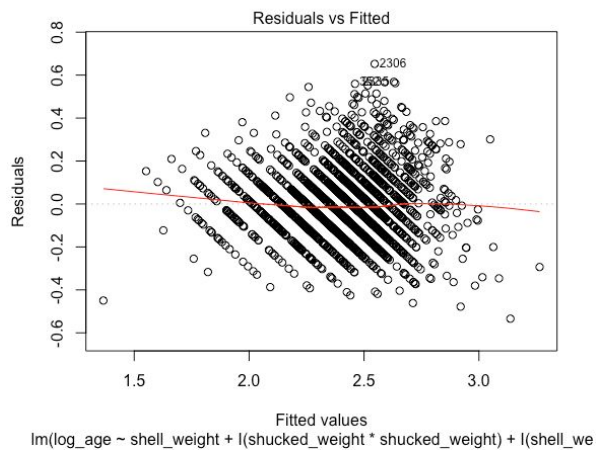
> anova(model_fit3)
Analysis of Variance Table

Response: log_age

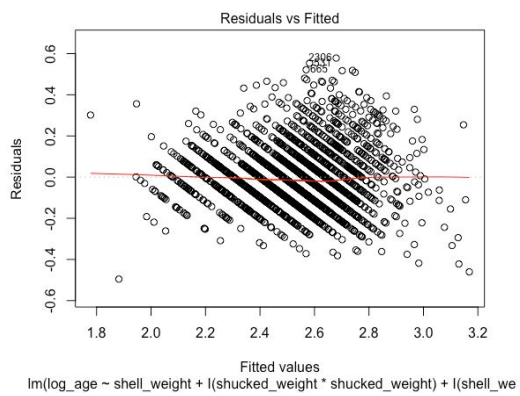
Df Sum Sq Mean Sq F value    Pr(>F)
shell_weight      1 49.802   49.802 2361.614 < 2.2e-16 ***
I(shucked_weight * shucked_weight) 1  9.366    9.366  444.143 < 2.2e-16 ***
I(shell_weight * shell_weight)      1  2.699    2.699 128.001 < 2.2e-16 ***
shucked_weight      1  3.820    3.820  181.142 < 2.2e-16 ***
sex                 2  1.440    0.720  34.140 2.582e-15 ***
whole_weight        1  1.045    1.045  49.546 2.623e-12 ***
length              1  0.738    0.738  34.974 3.898e-09 ***
height              1  0.482    0.482  22.862 1.863e-06 ***
shucked_weight:sex  2  1.760    0.880  41.728 < 2.2e-16 ***
shucked_weight:whole_weight 1  0.835    0.835  39.612 3.769e-10 ***
shell_weight:length  1  1.144    1.144  54.268 2.512e-13 ***
shell_weight:whole_weight 1  0.155    0.155   7.351 0.006758 **
Residuals        2074 43.737    0.021
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

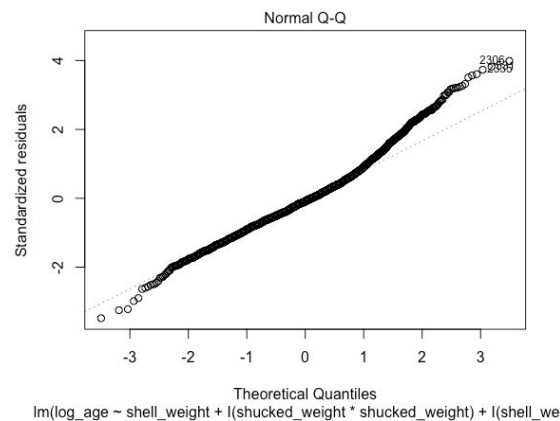
Plot 5e.)



Plot 5f.)







Plot 6a.)

```
> criterion_table
```

	p	cp	sse	mse	MSPE	aic	bic	Radj
model1	9	401.00892	63.89103	0.03071684	0.03073647	-7266.890	-7216.090	0.5770706
model2	24	43.27966	53.86986	0.02608710	0.02660001	-7593.290	-7457.823	0.6408159
model3	15	95.01421	55.67221	0.02684291	0.02691851	-7542.541	-7457.874	0.6304094

Plot 6b.)

(Intercept)	1.426616	shell_weight	4.185854	I(shucked_weight * shucked_weight)	3.408829
I(shell_weight * shell_weight)	-1.274031	shucked_weight	-1.621422	sexI	-0.273622
sexM	-0.117077	whole_weight	-0.544771	length	1.917785
height	6.885656	viscera_weight	-0.402232	shucked_weight:sexI	0.330107
shucked_weight:sexM	-0.003283	shucked_weight:whole_weight	-2.307068	shell_weight:length	-7.323386
shell_weight:whole_weight	-0.457542	sexI:height	0.878959	sexM:height	0.800871
shucked_weight:height	-12.451029	shell_weight:height	3.351807	length:height	-15.836451
whole_weight:length	2.394141	shell_weight:shucked_weight	3.854407	whole_weight:height	6.960776

```
> valid1
```

```
Call:
lm(formula = log_age ~ shell_weight + I(shucked_weight * shucked_weight) +
  I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +
  length + height + viscera_weight + shucked_weight:sex + shucked_weight:whole_weight +
  shell_weight:length + shell_weight:whole_weight + sex:height +
  shucked_weight:height + shell_weight:height + length:height +
  whole_weight:length + shell_weight:shucked_weight + whole_weight:height,
  data = valid)
```

Coefficients:

(Intercept)	1.311672	shell_weight	3.109796	I(shucked_weight * shucked_weight)	3.414739
I(shell_weight * shell_weight)	-1.375607	shucked_weight	-3.041453	sexI	-0.155613
sexM	-0.035010	whole_weight	0.547225	length	1.810303
height	7.966427	viscera_weight	-0.709109	shucked_weight:sexI	0.440129
shucked_weight:sexM	-0.001029	shucked_weight:whole_weight	-1.808878	shell_weight:length	-2.714643
shell_weight:whole_weight	1.063534	sexI:height	-0.127243	sexM:height	0.186155
shucked_weight:height	2.890567	shell_weight:height	-2.753230	length:height	-15.573027
whole_weight:length	1.140222	shell_weight:shucked_weight	-1.225042	whole_weight:height	0.988100

## Plot 7a.)

```
> summary(fitmodel)
```

Call:

```
lm(formula = log_age ~ shell_weight + I(shucked_weight * shucked_weight) +  
    I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +  
    length + height + viscera_weight + shucked_weight:sex + shucked_weight:whole_weight +  
    shell_weight:length + shell_weight:whole_weight + sex:height +  
    shucked_weight:height + shell_weight:height + length:height +  
    whole_weight:length + shell_weight:shucked_weight + whole_weight:height,  
    data = df_ab)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.79070	-0.09133	-0.01194	0.07585	0.67590

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.76251	0.06139	28.709	< 2e-16 ***
shell_weight	3.13647	0.63261	4.958	7.41e-07 ***
I(shucked_weight * shucked_weight)	2.92080	0.29854	9.784	< 2e-16 ***
I(shell_weight * shell_weight)	-0.45250	0.55768	-0.811	0.417189
shucked_weight	-1.73791	0.24744	-7.023	2.52e-12 ***
sexI	-0.22670	0.02998	-7.562	4.85e-14 ***
sexM	-0.09567	0.02788	-3.432	0.000605 ***
whole_weight	0.01321	0.21456	0.062	0.950923
length	1.39207	0.17405	7.998	1.63e-15 ***
height	5.00987	0.84093	5.958	2.77e-09 ***
viscera_weight	-0.49437	0.08683	-5.694	1.33e-08 ***
shucked_weight:sexI	0.28993	0.06834	4.242	2.26e-05 ***
shucked_weight:sexM	-0.02094	0.03826	-0.547	0.584195
shucked_weight:whole_weight	-1.65877	0.27455	-6.042	1.66e-09 ***
shell_weight:length	-4.49900	1.59659	-2.818	0.004857 **
shell_weight:whole_weight	-0.27245	0.45931	-0.593	0.553091
sexI:height	0.71965	0.30601	2.352	0.018733 *
sexM:height	0.67567	0.23924	2.824	0.004761 **
shucked_weight:height	-7.88024	2.21287	-3.561	0.000373 ***
shell_weight:height	0.44171	2.68347	0.165	0.869263
length:height	-11.06461	2.24489	-4.929	8.60e-07 ***
whole_weight:length	1.40608	0.44626	3.151	0.001640 **
shell_weight:shucked_weight	1.68886	0.71113	2.375	0.017598 *
whole_weight:height	4.69030	1.62815	2.881	0.003988 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1433 on 4153 degrees of freedom

Multiple R-squared: 0.6425, Adjusted R-squared: 0.6405

```
> anova(fitmodel)
```

Analysis of Variance Table

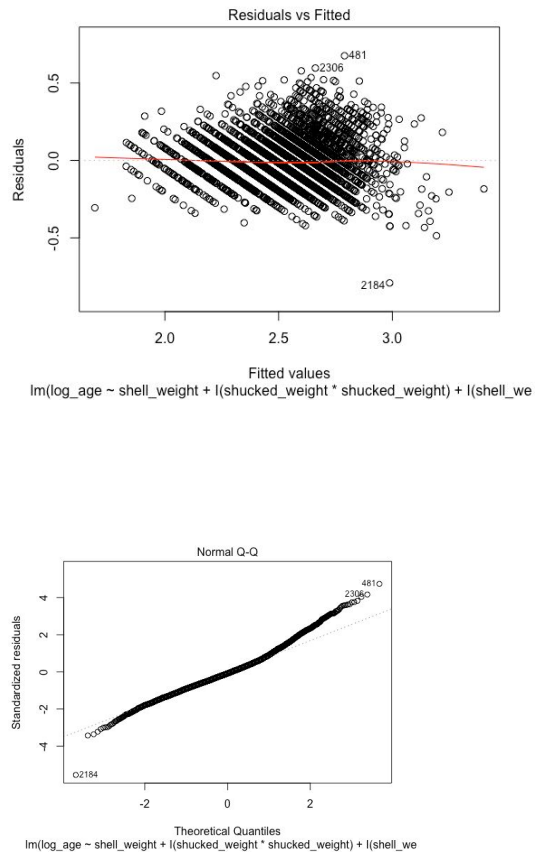
Response: log\_age

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
shell_weight	1	104.923	104.923	5110.6407	< 2.2e-16 ***
I(shucked_weight * shucked_weight)	1	17.499	17.499	852.3425	< 2.2e-16 ***
I(shell_weight * shell_weight)	1	6.271	6.271	305.4269	< 2.2e-16 ***
shucked_weight	1	6.229	6.229	303.3946	< 2.2e-16 ***
sex	2	2.382	1.191	58.0221	< 2.2e-16 ***
whole_weight	1	2.039	2.039	99.2990	< 2.2e-16 ***
length	1	2.032	2.032	98.9550	< 2.2e-16 ***
height	1	1.118	1.118	54.4548	1.911e-13 ***
viscera_weight	1	1.644	1.644	80.0708	< 2.2e-16 ***
shucked_weight:sex	2	3.364	1.682	81.9197	< 2.2e-16 ***
shucked_weight:whole_weight	1	1.815	1.815	88.4288	< 2.2e-16 ***
shell_weight:length	1	2.145	2.145	104.4728	< 2.2e-16 ***
shell_weight:whole_weight	1	0.272	0.272	13.2577	0.0002748 ***
sex:height	2	0.323	0.161	7.8555	0.0003934 ***
shucked_weight:height	1	0.372	0.372	18.1180	2.122e-05 ***
shell_weight:height	1	0.069	0.069	3.3553	0.0670610 .
length:height	1	0.158	0.158	7.7191	0.0054885 **
whole_weight:length	1	0.387	0.387	18.8413	1.454e-05 ***
shell_weight:shucked_weight	1	0.020	0.020	0.9960	0.3183498
whole_weight:height	1	0.170	0.170	8.2988	0.0039876 **
Residuals	4153	85.263	0.021		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Plot 7b.)



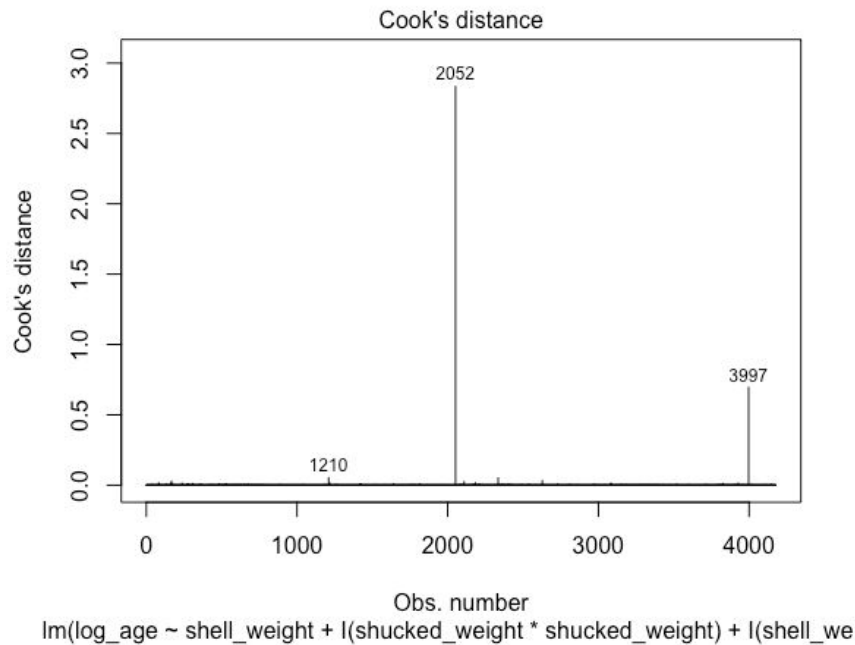
Plot 8a.)

```
> #X outliers
> h = as.vector(influence(fitmodel)$hat)
> index.X = which(h > (2*p/n))
> index.X
[1] 82 84 130 150 158 164 165 166 167 169 170 237 239 271 278 308 335 356 359 507
[21] 526 527 695 883 892 1000 1035 1052 1146 1175 1194 1202 1208 1210 1211 1217 1258 1265 1417 1418
[41] 1420 1427 1428 1429 1529 1638 1700 1749 1755 1762 1763 1764 1813 2052 2085 2091 2109 2158 2162 2170
[61] 2178 2180 2209 2210 2213 2335 2381 2382 2538 2545 2624 2625 2628 2642 2676 2708 2729 2791 2802 2811
[81] 2812 2855 2864 2971 3008 3009 3035 3051 3082 3087 3129 3150 3152 3189 3217 3319 3339 3389 3428 3519
[101] 3600 3714 3716 3733 3801 3815 3828 3861 3878 3903 3904 3929 3962 3963 3997 4018 4053 4149
```

Plot 8b.)

```
> head(sort(abs(stu.res.del), decreasing=TRUE))
2306 2335 352 2336 665 531
4.061266 3.856943 3.531064 3.502221 3.491077 3.481108
> qt(1-.1/(2*n), n-3-1) #Bonferroni's Threshold
[1] 4.074353
```

Plot 8c.)



Plot 9a.)

```
> summary(fitmodel)
```

```
Call:
lm(formula = log_age ~ shell_weight + I(shucked_weight * shucked_weight) +
  I(shell_weight * shell_weight) + shucked_weight + sex + whole_weight +
  length + height + viscera_weight + shucked_weight:sex + shucked_weight:whole_weight +
  shell_weight:length + shell_weight:whole_weight + sex:height +
  shucked_weight:height + shell_weight:height + length:height +
  whole_weight:length + shell_weight:shucked_weight + whole_weight:height,
  data = ab_clean)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.79132 -0.09114 -0.01215  0.07574  0.67040
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.66828    0.06794   24.557 < 2e-16 ***
shell_weight    3.20255    0.63223    5.065 4.25e-07 ***
I(shucked_weight * shucked_weight)  2.87382    0.29855    9.626 < 2e-16 ***
I(shell_weight * shell_weight)   -0.64043    0.56009   -1.143 0.252926
shucked_weight  -2.12216    0.27441   -7.733 1.30e-14 ***
sexI            -0.15470    0.03736   -4.141 3.53e-05 ***
sexM            -0.05188    0.03098   -1.675 0.094100 .
whole_weight     0.08285    0.21541    0.385 0.700528
length          1.36527    0.17405    7.844 5.51e-15 ***
height          6.56224    0.96826    6.777 1.40e-11 ***
viscera_weight  -0.50288    0.08677   -5.796 7.32e-09 ***
shucked_weight:sexI  0.39675    0.07588    5.228 1.79e-07 ***
shucked_weight:sexM  0.02044    0.04031    0.507 0.612125
shucked_weight:whole_weight -1.68586    0.27437   -6.144 8.78e-10 ***
shell_weight:length -4.52520    1.59481   -2.837 0.004570 **
shell_weight:whole_weight  0.09310    0.47260    0.197 0.843847
sexI:height      -0.04013    0.38600   -0.104 0.917198
sexM:height      0.27789    0.26895    1.033 0.301555
shucked_weight:height -4.04347    2.51052   -1.611 0.107341
shell_weight:height -0.10945    2.68589   -0.041 0.967496
length:height    -12.25895    2.27277   -5.394 7.28e-08 ***
whole_weight:length  1.60771    0.45013    3.572 0.000359 ***
shell_weight:shucked_weight  1.13739    0.73064    1.557 0.119615
whole_weight:height  2.82140    1.72658    1.634 0.102314
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1431 on 4152 degrees of freedom
Multiple R-squared:  0.6434,    Adjusted R-squared:  0.6414
```



```
> anova(fitmodel)
```

Analysis of Variance Table

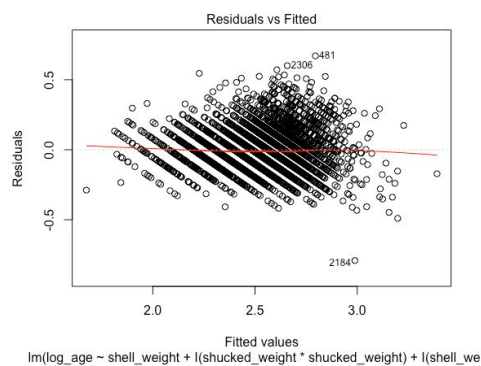
Response: log\_age

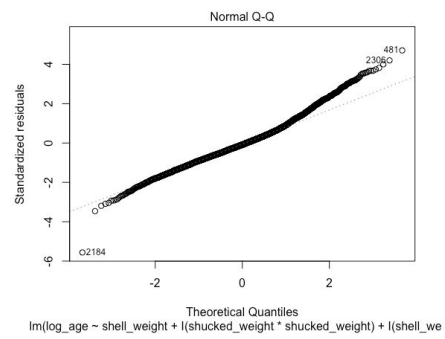
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
shell_weight	1	104.906	104.906	5121.3396	< 2.2e-16	***
I(shucked_weight * shucked_weight)	1	17.499	17.499	854.2712	< 2.2e-16	***
I(shell_weight * shell_weight)	1	6.270	6.270	306.1155	< 2.2e-16	***
shucked_weight	1	6.238	6.238	304.5315	< 2.2e-16	***
sex	2	2.377	1.189	58.0308	< 2.2e-16	***
whole_weight	1	2.041	2.041	99.6371	< 2.2e-16	***
length	1	2.035	2.035	99.3417	< 2.2e-16	***
height	1	1.983	1.983	96.8134	< 2.2e-16	***
viscera_weight	1	1.755	1.755	85.6618	< 2.2e-16	***
shucked_weight:sex	2	3.118	1.559	76.0970	< 2.2e-16	***
shucked_weight:whole_weight	1	1.672	1.672	81.6394	< 2.2e-16	***
shell_weight:length	1	1.884	1.884	91.9921	< 2.2e-16	***
shell_weight:whole_weight	1	0.237	0.237	11.5560	0.0006817	***
sex:height	2	0.110	0.055	2.6832	0.0684658	.
shucked_weight:height	1	0.510	0.510	24.8835	6.339e-07	***
shell_weight:height	1	0.003	0.003	0.1430	0.7052881	
length:height	1	0.276	0.276	13.4514	0.0002479	***
whole_weight:length	1	0.447	0.447	21.8063	3.111e-06	***
shell_weight:shucked_weight	1	0.012	0.012	0.5900	0.4424702	
whole_weight:height	1	0.055	0.055	2.6703	0.1023143	
Residuals	4152	85.050	0.020			

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Plot 9b.)





3. Code:

## References

1.

Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn and Wes B Ford (1994)  
"The Population Biology of Abalone (*Haliotis* species) in Tasmania. I. Blacklip Abalone (*H. rubra*) from the North Coast and Islands of Bass Strait", Sea Fisheries Division, Technical Report No. 48 .

2.

Data Source

UCI machine learning Repository.