Homework 2

Jules Merigot (8488256)

October 04, 2022

PSTAT 131/231 Statistical Machine Learning - Fall 2022

Linear Regression

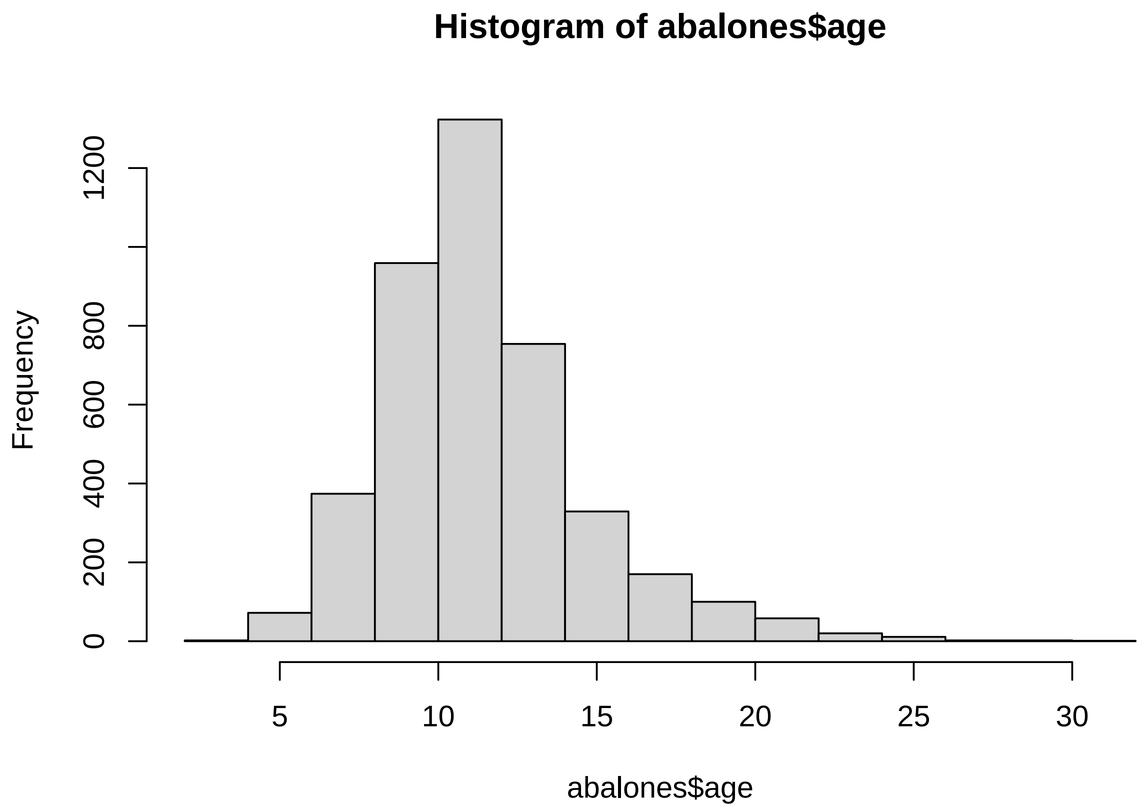
Question 1

```
# Adding the age variable as a column to the Abalone data frame
abalones <- aba_data %>%
  mutate(age = rings + 1.5)

# Checking to see it was correctly added
head(abalones)
```

```
##   type longest_shell diameter height whole_weight shucked_weight viscera_weight
## 1    M         0.455    0.365  0.095       0.5140         0.2245         0.1010
## 2    M         0.350    0.265  0.090       0.2255         0.0995         0.0485
## 3    F         0.530    0.420  0.135       0.6770         0.2565         0.1415
## 4    M         0.440    0.365  0.125       0.5160         0.2155         0.1140
## 5    I         0.330    0.255  0.080       0.2050         0.0895         0.0395
## 6    I         0.425    0.300  0.095       0.3515         0.1410         0.0775
##   shell_weight rings  age
## 1        0.150    15 16.5
## 2        0.070     7  8.5
## 3        0.210     9 10.5
## 4        0.155    10 11.5
## 5        0.055     7  8.5
## 6        0.120     8  9.5
```

```
# Making a histogram of the age in order to asses the distribution
hist(abalones$age)
```



Using a histogram, we can see that the age of the abalones is normally distributed and skewed right, with an average age of about 11 years old. While there are more outliers that are older in age, most abalones tend to live between 5 and 15 years.

Question 2

```
set.seed(8488)

abalones_split <- initial_split(abalones, prop=0.80, strata=age)

abalones_train <- training(abalones_split)
abalones_test <- testing(abalones_split)
```

Question 3

I shouldn't include *rings* in the recipe to predict *age* since the *age* variable was calculated and added to the abalones dataset using the *rings* variable. As done in Question 1, we copied the *rings* column and added 1.5 in order to create the *age* column.

```
abalones_recipe <- recipe(age ~ ., data=abalones_train) %>%
  step_rm(rings) %>% # removing rings variable
  step_dummy(all_nominal_predictors()) %>% # Step 1: dummy code categorical predictors
```

```r
  step_interact(terms = ~ type_M:shucked_weight) %>% # Step 2: creating interactions
  step_interact(terms = ~ type_I:shucked_weight) %>%
  step_interact(terms = ~ longest_shell:diameter) %>%
  step_interact(terms = ~ shucked_weight:shell_weight) %>%
  step_center(all_predictors()) %>% # Step 3: centering predictors
  step_scale(all_predictors()) # Step 4: scaling predictors
abalones_recipe
```

Question 4

```r
lm_model <- linear_reg() %>%
  set_engine("lm")
```

Question 5

```r
lm_wflow <- workflow() %>%
  add_model(lm_model) %>%
  add_recipe(abalones_recipe)
```

Question 6

```r
lm_fit <- fit(lm_wflow, abalones_train)

lm_fit %>%
  # This returns the parsnip object:
  extract_fit_parsnip() %>%
  # Now tidy the linear model object:
  tidy()
```

```
## # A tibble: 14 x 5
##    term                         estimate std.error statistic  p.value
##    <chr>                           <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                   11.4       0.0370   309.    0
##  2 longest_shell                  0.448     0.283      1.59  1.13e- 1
##  3 diameter                       2.10      0.311      6.77  1.53e-11
##  4 height                         0.273     0.0691     3.96  7.71e- 5
##  5 whole_weight                   5.27      0.395     13.3   1.50e-39
##  6 shucked_weight                -4.56      0.256    -17.8   6.11e-68
##  7 viscera_weight                -0.942     0.158     -5.97  2.65e- 9
##  8 shell_weight                   1.50      0.212      7.08  1.74e-12
##  9 type_I                        -1.04      0.116     -8.99  4.18e-19
## 10 type_M                        -0.297     0.103     -2.89  3.89e- 3
## 11 type_M_x_shucked_weight        0.371     0.109      3.41  6.51e- 4
## 12 type_I_x_shucked_weight        0.602     0.0871     6.91  5.73e-12
## 13 longest_shell_x_diameter      -2.80      0.397     -7.05  2.21e-12
## 14 shucked_weight_x_shell_weight -0.0378    0.200     -0.189 8.50e- 1
```