Homework2

Yingshan Li (7937790)

April 06, 2022

```
abalone <- read.csv(file = "abalone.csv")
```

Question 1

```
abalone2 <- abalone %>%
  mutate(age = abalone$rings + 1.5)
```

Question 2

```
set.seed(2231)

abalone_split <- initial_split(abalone2, prop = 0.80, strata = age)

abalone_train <- training(abalone_split)
abalone_test <- testing(abalone_split)
```

Question 3

```
abalone_recipe <- recipe(age ~ type + longest_shell + diameter + height + whole_weight + shucked_weight
  step_dummy(all_nominal_predictors()) %>%
  step_interact(~ starts_with("type"):shucked_weight + longest_shell:diameter + shucked_weight:shell_we
  step_center(all_predictors()) %>%
  step_scale(all_predictors())
```

Question 4

```
lm_mod <- linear_reg() %>%
  set_engine("lm")
```

Question 5

```
lm_wflow <- workflow() %>%
  add_model(lm_mod) %>%
  add_recipe(abalone_recipe)
```

Question 6 fit the linear model

```
lm_fit <- fit(lm_wflow, abalone_train)
```

```
lm_fit %>%
  extract_fit_parsnip() %>%
  tidy()
```

```
## # A tibble: 14 x 5
##    term                          estimate std.error statistic  p.value
##    <chr>                            <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                      11.4      0.0372   307.    0
```

```
##  2 longest_shell                   0.679     0.286      2.37  1.77e- 2
##  3 diameter                        1.87      0.309      6.04  1.73e- 9
##  4 height                          0.236     0.0689     3.42  6.26e- 4
##  5 whole_weight                    5.55      0.417     13.3   2.27e-39
##  6 shucked_weight                 -4.71      0.262    -18.0   5.50e-69
##  7 viscera_weight                 -0.971     0.160     -6.08  1.36e- 9
##  8 shell_weight                    1.19      0.223      5.34  1.01e- 7
##  9 type_I                         -0.992     0.115     -8.60  1.22e-17
## 10 type_M                         -0.284     0.102     -2.77  5.62e- 3
## 11 type_I_x_shucked_weight         0.522     0.0869     6.00  2.20e- 9
## 12 type_M_x_shucked_weight         0.309     0.108      2.86  4.31e- 3
## 13 longest_shell_x_diameter       -2.59      0.398     -6.49  9.66e-11
## 14 shucked_weight_x_shell_weight  -0.0223    0.204     -0.109 9.13e- 1
```

```
Preicted_age <-predict(lm_fit, data.frame(type = "F", longest_shell = 0.50, diameter = 0.10, height = 0
Preicted_age
```

```
## # A tibble: 1 x 1
##   .pred
##   <dbl>
## 1  24.1
```

Question 7

```
abalone_metrics <- metric_set(rmse, rsq, mae)
```

```
abalone_train_res <- predict(lm_fit, new_data = abalone_train %>% select(-age, -rings))
abalone_train_res %>%
  head()
```

```
## # A tibble: 6 x 1
##   .pred
##   <dbl>
## 1  8.04
## 2  9.34
## 3  9.64
## 4 10.4
## 5  5.75
## 6  5.91
```

```
abalone_train_res <- bind_cols(abalone_train_res, abalone_train %>% select(age))
abalone_train_res %>%
  head()
```

```
## # A tibble: 6 x 2
##   .pred   age
##   <dbl> <dbl>
## 1  8.04   8.5
## 2  9.34   9.5
## 3  9.64   8.5
## 4 10.4    8.5
## 5  5.75   6.5
## 6  5.91   5.5
```

```
abalone_metrics(abalone_train_res, truth = age, estimate = .pred)
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
```

```
##   <chr>   <chr>          <dbl>
## 1 rmse    standard       2.15
## 2 rsq     standard       0.553
## 3 mae     standard       1.54
```

R^2 : 0.553. 55.3% of the variation in the response can be explained by the predictors.