Homework 2

Kyle Kim

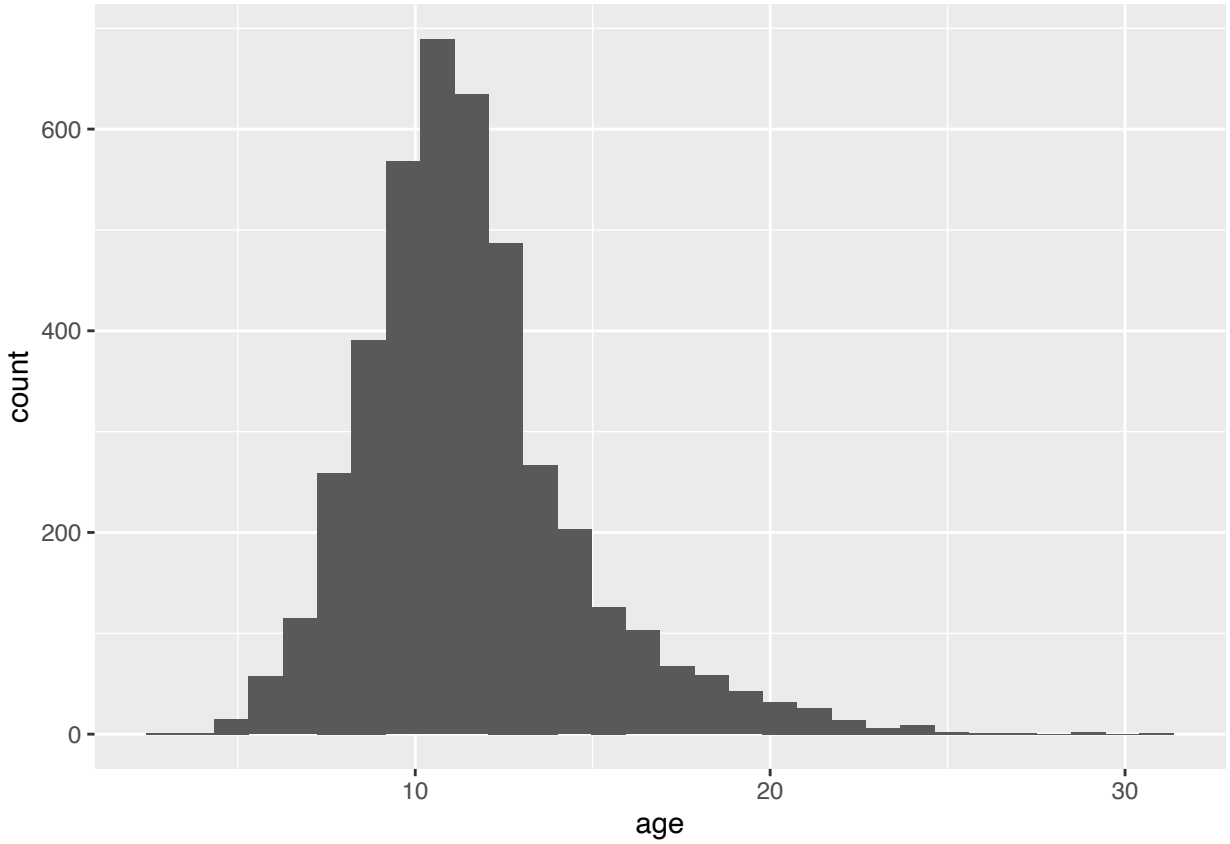
2022-10-16

Q1

```
abal <- abal %>%
  mutate(age = rings + 1.5)
abal
```

```
## # A tibble: 4,177 x 10
##    type  longest_sh~1 diame~2 height whole~3 shuck~4 visce~5 shell~6 rings   age
##    <chr>        <dbl>   <dbl>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl>
##  1 M            0.455   0.365  0.095   0.514  0.224   0.101    0.15     15  16.5
##  2 M            0.35    0.265  0.09    0.226  0.0995  0.0485   0.07      7   8.5
##  3 F            0.53    0.42   0.135   0.677  0.256   0.142    0.21      9  10.5
##  4 M            0.44    0.365  0.125   0.516  0.216   0.114    0.155    10  11.5
##  5 I            0.33    0.255  0.08    0.205  0.0895  0.0395   0.055     7   8.5
##  6 I            0.425   0.3    0.095   0.352  0.141   0.0775   0.12      8   9.5
##  7 F            0.53    0.415  0.15    0.778  0.237   0.142    0.33     20  21.5
##  8 F            0.545   0.425  0.125   0.768  0.294   0.150    0.26     16  17.5
##  9 M            0.475   0.37   0.125   0.509  0.216   0.112    0.165     9  10.5
## 10 F            0.55    0.44   0.15    0.894  0.314   0.151    0.32     19  20.5
## # ... with 4,167 more rows, and abbreviated variable names 1: longest_shell,
## #   2: diameter, 3: whole_weight, 4: shucked_weight, 5: viscera_weight,
## #   6: shell_weight
```

```
abal %>%
  ggplot(aes(x=age)) + geom_histogram()
```




```
## The graph is normally distributed, with the right skew being the longer end.
```

Q2

```
set.seed(1000)
splitdata <- abal %>% initial_split(prop = 0.8, strata = age)
abal_training <- training(splitdata)  # 80%, Tibble: 3,340 x 10
abal_testing  <- testing(splitdata)  # 20%, Tibble: 837 x 10
```

Q3

```
recipe <- recipe(age ~ ., abal_training) %>%
  step_rm(rings) %>%  # Remove 'rings' because not a predictor of age
  step_dummy(all_nominal_predictors()) %>% # step 1
  step_interact(terms = ~ starts_with("type"):shucked_weight + longest_shell:diameter + shucked_weight:s
  step_normalize(all_predictors()) # step 3 & 4

recipe
```

```
## Recipe
##
```

```
## Inputs:
##
##       role #variables
##    outcome          1
##  predictor          9
##
## Operations:
##
## Variables removed rings
## Dummy variables from all_nominal_predictors()
## Interactions with starts_with("type"):shucked_weight + longest_shell...
## Centering and scaling for all_predictors()
```

Q4

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

```
## Linear Regression Model Specification (regression)
##
## Computational engine: lm
```

Q5

```
wrkflow <- workflow() %>%
  add_model(lm) %>%
  add_recipe(recipe)

fitted_model <- fit(wrkflow, abal_training)
fitted_model
```

```
## == Workflow [trained] ==========================================================
## Preprocessor: Recipe
## Model: linear_reg()
##
## -- Preprocessor ----------------------------------------------------------------
## 4 Recipe Steps
##
## * step_rm()
## * step_dummy()
## * step_interact()
## * step_normalize()
##
## -- Model -----------------------------------------------------------------------
##
## Call:
## stats::lm(formula = ..y ~ ., data = data)
##
```

```
## Coefficients:
##                     (Intercept)                   longest_shell
##                         11.4365                          0.3123
##                        diameter                          height
##                          2.2915                          0.2042
##                    whole_weight                  shucked_weight
##                          4.8368                         -4.2184
##                  viscera_weight                    shell_weight
##                         -0.9285                          1.7573
##                          type_I                          type_M
##                         -0.9092                         -0.2428
##       type_I_x_shucked_weight        type_M_x_shucked_weight
##                          0.4619                          0.2598
##      longest_shell_x_diameter  shucked_weight_x_shell_weight
##                         -2.7793                         -0.1637
```

Q6

```
prediction <- tibble(type = 'F', longest_shell = 0.50, diameter = 0.10, height = 0.30, whole_weight = 4

femaleabal <- predict(fitted_model, prediction)
femaleabal
```

```
## # A tibble: 1 x 1
##   .pred
##   <dbl>
## 1  22.7
```

```
# The predicted age of a female abalone is approximately 22.68 years.
```

Q7

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

predvalue_tibble <- predict(fitted_model, abal_training) %>%
  bind_cols(abal_training %>% select(age))

mset(predvalue_tibble, age, .pred)
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rsq     standard       0.562
## 2 rmse    standard       2.15
## 3 mae     standard       1.55
```

```
# Our R^2 value is low at approximately 0.562, meaning that approx. 56% of the
# variability is explained by the model.
```