

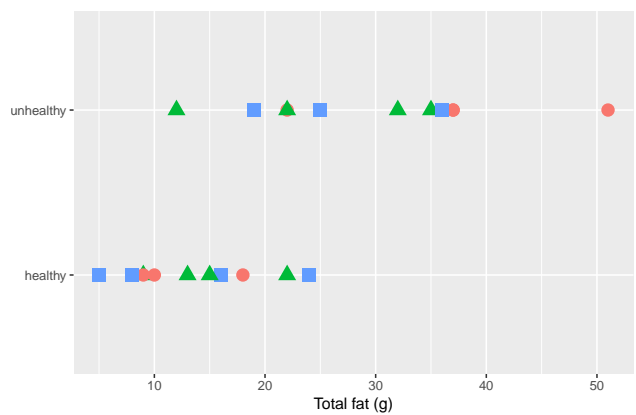
# WU #11: Math 150, Spring 2023

Thursday 2/23/2023

Name: \_\_\_\_\_

Names of people you worked with: \_\_\_\_\_

Consider the following dataset measuring fat content and calories for 12 fast food items.<sup>1</sup> Cross validated models have been fit for  $v = 3$  folds. Unhealthy are items with more than 400 calories; healthy items have 400 calories or less.



The values of the observations in group b are as follows:

```
## # A tibble: 8 x 3
##   type      total_fat fold
##   <chr>      <dbl> <chr>
## 1 healthy         9 b
## 2 healthy        13 b
## 3 healthy        15 b
## 4 healthy        22 b
## 5 unhealthy      12 b
## 6 unhealthy      32 b
## 7 unhealthy      22 b
## 8 unhealthy      35 b
```

Calculate accuracy for the observations in fold b. Note that “success” is “unhealthy”.

## a and b points

```
## # A tibble: 2 x 5
##   term      estimate std.error statistic p.value
##   <chr>      <dbl>    <dbl>    <dbl>  <dbl>
## 1 (Intercept) -4.97      2.46     -2.02  0.0431
## 2 total_fat    0.253    0.128      1.97  0.0485
```

## a and c points

```
## # A tibble: 2 x 5
##   term      estimate std.error statistic p.value
##   <chr>      <dbl>    <dbl>    <dbl>  <dbl>
## 1 (Intercept) -7.97      4.84     -1.65  0.0994
## 2 total_fat    0.391    0.230      1.70  0.0899
```

## b and c points

```
## # A tibble: 2 x 5
##   term      estimate std.error statistic p.value
##   <chr>      <dbl>    <dbl>    <dbl>  <dbl>
## 1 (Intercept) -4.23      2.14     -1.98  0.0482
## 2 total_fat    0.216    0.105      2.05  0.0400
```

<sup>1</sup>the data actually come from a much larger and real dataset

## Solution:

```
b_pts <- ff %>%
  filter(fold == "b")

ac_mod <- ff %>%
  filter(fold != "b") %>%
  glm(as.factor(type) ~ total_fat, data = ., family = "binomial")

ac_mod %>% tidy()

## # A tibble: 2 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>  <dbl>
## 1 (Intercept)   -7.97     4.84     -1.65  0.0994
## 2 total_fat     0.391    0.230     1.70  0.0899

preds <- ac_mod %>%
  augment(newdata = b_pts, type.predict = "response") %>%
  mutate(pred_class = ifelse(.fitted > .5, "unhealthy", "healthy"))

preds %>% select(type, pred_class) %>% table

##           pred_class
## type      healthy unhealthy
## healthy         3         1
## unhealthy       1         3
```

Accuracy is  $6/8 = 0.75$