Lab 14 - Logistic Regression

Who Are You?

Loading Packages

Run the code chunk below to load packages needed for this lab.

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Crab Species Identification

We have a data set about Leptograpsus crabs. There are two species of this crab; we will attempt to predict the species of a crab based on measurements of its physical dimensions. The data we are working with contains 5 morphological measurements on 50 crabs each of two species and both sexes, of Leptograpsus crabs collected at Fremantle, W. Australia.

Our data were original presented in

Campbell, N.A. and Mahon, R.J. (1974) A multivariate study of variation in two species of rock crab of genus Leptograpsus. Australian Journal of Zoology 22, 417–425.

They have been discussed previously in

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

They are available from the MASS package for R:

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
crabs <- crabs %>%
  mutate(
    sp_0_1 = as.numeric(sp) - 1
  )

head(crabs)
```

```
##   sp sex index   FL  RW   CL   CW  BD sp_0_1
## 1  B   M     1  8.1 6.7 16.1 19.0 7.0      0
## 2  B   M     2  8.8 7.7 18.1 20.8 7.4      0
## 3  B   M     3  9.2 7.8 19.0 22.4 7.7      0
## 4  B   M     4  9.6 7.9 20.1 23.1 8.2      0
## 5  B   M     5  9.8 8.0 20.3 23.0 8.2      0
## 6  B   M     6 10.8 9.0 23.0 26.5 9.8      0
```

```
dim(crabs)
```

```
## [1] 200   9
```

The variables in this data set are as follows:

- `sp`: species - "B" or "O" for blue or orange.
- `sex`: the crab's sex
- `index`: index 1:50 within each of the four groups.
- `FL`: frontal lobe size (mm).
- `RW`: rear width (mm).
- `CL`: carapace length (mm).
- `CW`: carapace width (mm).
- `BD`: body depth (mm).
- `sp_0_1`: species encoded as a binary variable. 0 represents species "B" and 1 represents species "O"

For today, let's explore prediction of the `sp_0_1` variable using `FL` as the explanatory variable.

Problem 1: Make a scatter plot of the data with `sp_0_1` on the vertical axis and `FL` on the horizontal axis.

Problem 2: Fit a logistic regression model to the training set crabs with `sp_0_1` as the response and `FL` as the explanatory variable. Print out a summary of the fit.

Problem 3: Make a scatter plot of the training set data with the estimated logistic regression curve overlaid on top.

Problem 4: What is the interpretation of the estimated coefficient for `FL` in your model?

Problem 5: If you found a new crab and you weren't sure which species it was, but it had a frontal lobe size of 19 mm, what would you estimate its probability of being an orange crab to be, based on this model? What would the estimated odds that the crab is an orange crab be?

Problem 6: Repeat problem 5, but now your crab has a frontal lobe size of 20 mm.

Problem 7: By comparing the estimated odds of being an orange crab from problems 5 and 6, verify that your interpretation in problem 4 was correct.