

Class Activity: Logistic Regression with a Single Explanatory Variable

STAT 340: Applied Regression

Exercise: Moth Coloration and Natural Selection

In 1972, J. A. Bishop published data from a study on moths in a cline (“a region where two color morphs of one species exist on opposite ends of the environmental gradient”) near Liverpool, England. In this particular cline, some moths are dark morph, and some are a lighter, “pepper-and-salt” morph. Because Liverpool had a large number of mills, which produced smoke that blackened the trees, dark morph moths are plentiful in the area; the lighter morph moths are more prevalent elsewhere. To study natural selection, seven locations, which were at varying distances from Liverpool, were selected. In each location, eight trees were chosen at random, and equal numbers of dead (frozen) light and dark moths were glued to the trees in lifelike positions. Twenty-four hours later, the number of removed moths from each morph was recorded; moths were assumed to be removed by predators.

The data from this study are available in the `Sleuth3` package in R (`case2102`).

```
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(tidyr)
library(Sleuth3)

moths <- case2102
```

The variables are as follows:

- **Morph:** a factor with two levels denoting the morph, `light` and `dark`
- **Distance:** distance from Liverpool, England in kilometers (km)
- **Placed:** number of moths placed
- **Removed:** number of moths removed

Question of interest: Does the probability of being removed (between light and dark morph moths) differ, and does the proportion removed depend on distance?

The GLM

The response variable here is

$$Y_i = \begin{cases} 1 & \text{if moth number } i \text{ is removed} \\ 0 & \text{otherwise} \end{cases}.$$

(a) What is the random component of the model?

(b) Consider the question of interest from above. Write down the linear component of the GLM such that it addresses this question (it should include the appropriate explanatory variable(s)). For this exercise, you do not need to use matrix form, but you are welcome to if you want to practice.

(c) What is the link function?

Fitting the GLM

We will assume for now that all the assumptions are met for the logistic regression model. We will discuss these more in the coming classes.

```
## Require 0/1 as response for glm
moths <- moths %>%
  uncount(Placed)

bernResponse <- NULL
for (i in 1:nrow(case2102)){
  bernResponse <- c(bernResponse,
                    c(rep(1, case2102$Removed[i]), rep(0, case2102$Placed[i]-case2102$Removed[i])))
}
moths$bernResponse <- bernResponse

## Fit logistic regression model
mod1 <- glm(bernResponse ~ Distance + Morph + Distance:Morph, data=moths, family=binomial)
summary(mod1)
```

```
##
## Call:
## glm(formula = bernResponse ~ Distance + Morph + Distance:Morph,
##      family = binomial, data = moths)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1013  -0.8914  -0.7727   1.3341   1.7077
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.128987   0.197906  -5.705 1.17e-08 ***
## Distance       0.018502   0.005645   3.277 0.001048 **
## Morphlight     0.411257   0.274490   1.498 0.134065
## Distance:Morphlight -0.027789  0.008085  -3.437 0.000588 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1207.9  on 967  degrees of freedom
## Residual deviance: 1185.7  on 964  degrees of freedom
## AIC: 1193.7
##
## Number of Fisher Scoring iterations: 4
```

- (a) What is the estimated probability that a moth of light morph located 30 km from Liverpool is removed by a predator?
- (b) What is the interpretation of $\hat{\beta}_{\text{Distance}}$, the coefficient for distance in terms of the *log odds* of removal?
- (c) What is the interpretation of $\hat{\beta}_{\text{Distance}}$, the coefficient for distance in terms of the *odds* of removal?
- (d) Is there evidence that the effect of distance on the odds of removal is different for moths of light morph versus moths of dark morph?