## Week #11: Logistic Regression Models

Cheng Peng

4/9/2021

## Contents

1 Introduction 1

2 Motivational Examples and Practical Question

1

## 1 Introduction

Linear regression models are used to assess the association between the continuous response variable and other predictor variables. If the response variable is a binary categorical variable, the linear regression model is not appropriate. We need to a new model, logistic regression models, to assess the association the the binary rrsponse variable and other predictor variables.

This module focuses on the regression model with a binary response.

## 2 Motivational Examples and Practical Question

**Example 1.** Suzuki et al. (2006) measured sand grain size on 28 beaches in Japan and observed the presence or absence of the burrowing wolf spider Lycosa ishikariana on each beach. Sand grain size is a measurement variable, and spider presence or absence is a nominal variable. Spider presence or absence is the dependent variable; if there is a relationship between the two variables, it would be sand grain size affecting spiders, not the presence of spiders affecting the sand.

| Grain Size |         | Numerical | Grain Size | 6236 8000 | Numerical |
|------------|---------|-----------|------------|-----------|-----------|
| (mm)       | Status  | Status    | (mm)       | status    | Status    |
| 0.245      | absent  | 0         | 0.432      | absent    | 0         |
| 0.247      | absent  | 0         | 0.473      | present   | 1         |
| 0.285      | present | 1         | 0.509      | present   | 1         |
| 0.299      | present | 1         | 0.529      | present   | 1         |
| 0.327      | present | 1         | 0.561      | absent    | 0         |
| 0.347      | present | 1         | 0.569      | absent    | 0         |
| 0.356      | absent  | 0         | 0.594      | present   | 1         |
| 0.36       | present | 1         | 0.638      | present   | 1         |
| 0.363      | absent  | 0         | 0.656      | present   | 1         |
| 0.364      | present | 1         | 0.816      | present   | 1         |
| 0.398      | absent  | 0         | 0.853      | present   | 1         |
| 0.4        | present | 1         | 0.938      | present   | 1         |
| 0.409      | absent  | 0         | 1.036      | present   | 1         |
| 0.421      | present | 1         | 1.045      | present   | 1         |

Figure 1: Spider Data Table

```
grainsize=c(0.245, 0.247, 0.285, 0.299, 0.327, 0.347, 0.356, 0.360, 0.363, 0.364, 0.398, 0.400, 0.409, 0.421, 0.432, 0.473, 0.509, 0.529, 0.561, 0.569, 0.594, 0.638,
```

```
0.656, 0.816, 0.853, 0.938, 1.036, 1.045)

status=c(0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1)

spider = as.data.frame(cbind(grainsize = grainsize, status = status))
```

As usual, there are two objectives for this study.

- Association analysis: to determine whether there was a relationship between sand grain size and the presence or absence of the species.
- **Prediction analysis**: to predict the probability of a wolf spider population surviving on a beach with a particular sand grain size, to help determine which beaches to reintroduce the spider to.

**Example 2.**: Longevity in male fruit flies is positively associated with adult size. But reproduction has a high physiological cost that could impact longevity. A study looks at the association between longevity and adult size in male fruit flies kept under one of two conditions. One group is kept with sexually active females over the male's life span. The other group is cared for in the same way but kept with females that are not sexually active.

| Longevity | ThxLength | IndReprod | Longevity | ThxLength | IndReprod |
|-----------|-----------|-----------|-----------|-----------|-----------|
| 34        | 0.78      | 1         | 46        | 0.84      | 1         |
| 42        | 0.76      | 1         | 56        | 0.76      | 0         |
| 30        | 0.8       | 1         | 76        | 0.92      | 0         |
| 46        | 0.88      | 1         | 65        | 0.8       | 0         |
| 40        | 0.82      | 1         | 42        | 0.76      | 1         |
| 49        | 0.68      | 0         | 19        | 0.64      | 1         |
| 56        | 0.8       | 0         | 19        | 0.68      | 1         |
| 70        | 0.88      | 0         | 70        | 0.88      | 0         |
| 64        | 0.76      | 0         | 26        | 0.8       | 1         |
| 54        | 0.88      | 1         | 64        | 0.72      | 0         |
| 85        | 0.84      | 0         | 76        | 0.92      | 0         |
| 76        | 0.84      | 0         | 33        | 0.72      | 1         |
| 54        | 0.88      | 1         | 81        | 0.84      | 0         |
| 61        | 0.88      | 1         | 34        | 0.72      | 1         |
| 56        | 0.88      | 1         | 54        | 0.82      | 1         |
| 46        | 0.76      | 0         | 65        | 0.84      | 0         |
| 44        | 0.92      | 1         | 37        | 0.68      | 0         |
| 76        | 0.94      | 0         | 39        | 0.76      | 0         |
| 70        | 0.84      | 0         | 35        | 0.84      | 1         |
| 64        | 0.84      | 0         | 35        | 0.64      | 0         |
| 70        | 0.8       | 0         | 30        | 0.76      | 1         |
| 35        | 0.84      | 1         | 46        | 0.84      | 1         |
| 65        | 0.76      | 0         | 16        | 0.64      | 1         |
| 34        | 0.74      | 1         | 46        | 0.72      | 0         |

Figure 2: Fruit Flies Data Table

The above tale gives the longevity in days and thorax length in millimeters (mm) for the male fruit flies given an opportunity to reproduce (IndReprod = 1) and for those deprived of the opportunity (IndReprod = 0).1 Is there evidence that reproduction impacts male fruit fly longevity?