

Homework 3

Code:

```
# Load necessary libraries
install.packages("popbio")
library(popbio)

# Load the dataset
data(puffinbill)

# Display column names and a preview of the dataset
names(puffinbill)
print(puffinbill)

# Assign features to variables
sex <- puffinbill$sex
curlen <- puffinbill$curlen

# Apply encoding on the categorical column 'sex'
sexcode <- ifelse(sex == "F", 1, 0)
print(sexcode)

# Plot the observations
plot(curlen, jitter(sexcode, 0.15), pch = 19, xlab = "Bill Length (mm)", ylab = "Sex (0 - male, 1 - female)")

# Fit a generalized logistic regression model
model <- glm(sexcode ~ curlen, binomial)
summary(model)

# Explore the range of 'curlen' values
xv <- seq(min(curlen), max(curlen), 0.01)
summary(xv)

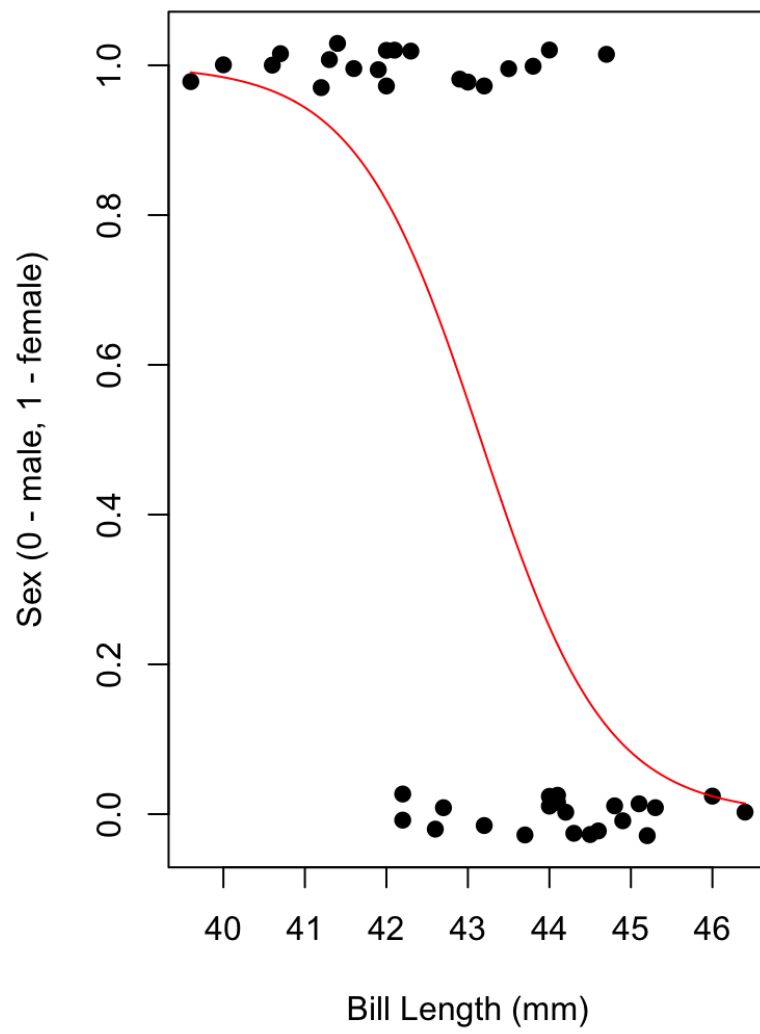
# Predict probabilities using the model
yv <- predict(model, list(curlen = xv), type = "response")
summary(yv)

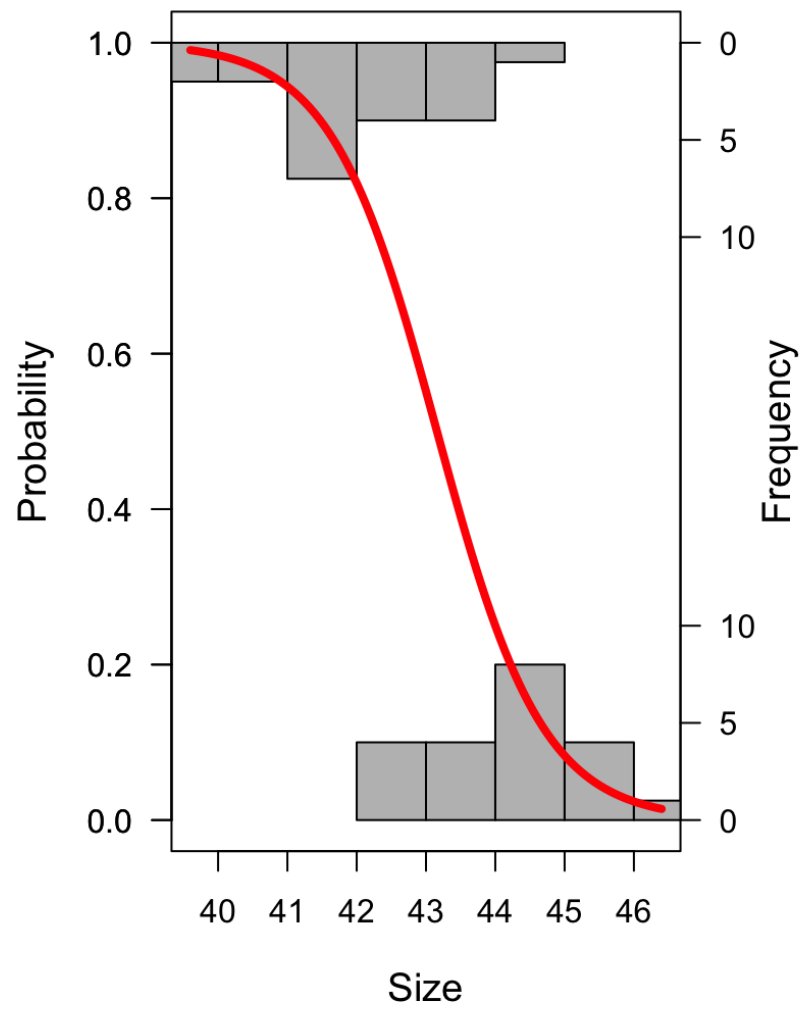
# Plot the logistic regression curve
lines(xv, yv, col = "red")

# Display a frequency distribution of each category
logi.hist.plot(curlen, sexcode, boxp = FALSE, type = "count", col = "gray", xlab = "Size")
```

Results:

Plots:





Console:

```
> names(puffinbill)
```

```
[1] "bird" "sex" "curlen"
```

```
> print(puffinbill)
```

```
bird sex curlen
1  1  M  44.1
2  2  M  44.0
3  3  M  44.1
4  4  M  42.6
5  5  M  46.4
6  6  M  42.2
7  7  M  44.5
8  8  M  44.2
9  9  M  43.7
10 10  M  44.0
11 11  M  45.3
12 12  M  44.3
13 13  M  46.0
14 14  M  45.2
15 15  M  43.2
16 16  M  45.1
17 17  M  44.6
18 18  M  42.2
19 19  M  42.7
20 20  M  44.9
21 21  M  44.8
22 22  F  43.0
23 23  F  42.0
24 24  F  43.2
25 25  F  42.9
26 26  F  40.0
27 27  F  40.6
28 28  F  42.1
29 29  F  42.0
30 30  F  43.5
31 31  F  41.4
32 32  F  41.6
33 33  F  44.7
34 34  F  41.3
35 35  F  40.7
36 36  F  41.2
37 37  F  42.3
38 38  F  44.0
39 39  F  39.6
40 40  F  43.8
41 41  F  41.9
```

```
> # Explore the range of 'curlen' values
> xv <- seq(min(curlen), max(curlen), 0.01)
> summary(xv)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 39.6   41.3   43.0   43.0   44.7   46.4
> # Predict probabilities using the model
> yv <- predict(model, list(curlen = xv), type = "response")
> summary(yv)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.01433 0.11795 0.55160 0.52276 0.91880 0.99048
> # Plot the logistic regression curve
> lines(xv, yv, col = "red")
```

```
> # Display a frequency distribution of each category  
> logi.hist.plot(curlen, sexcode, boxp = FALSE, type = "count", col = "gray", xlabel = "Size")  
>
```