Solution of assignment 6, ST2304

Problem 1 1. The inverse of the logit function is the logistic function. If logit(p) = z, then

$$p = \frac{1}{1 + e^{-\eta}} \tag{1}$$

in this case

$$\eta = \beta_0 + \beta_{\text{age}}\text{age} + \beta_{\log_{10}\text{ab}} \log_{10}\text{ab} \tag{2}$$

which means that we get

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_{\text{age}}\text{age} + \beta_{\log_{10}\text{ab}} \log_{10}\text{ab})}} \tag{3}$$

We set in for age =15 and antibody level =1000 and β's from the summary() of the logistic regression:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_{\text{age}}\text{age} + \beta_{\log_{10}\text{ab}} \log_{10}\text{ab})}} \tag{4}$$

The estimated probability of developing malaria is then: 0.04216440

```
> summary(malreg)

Call:
glm(formula = mal ~ age + log10(ab), family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8492  -0.7536  -0.4838   0.8809   2.5796

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.57234    0.95184   2.702 0.006883 **
age         -0.06546    0.06772  -0.967 0.333703
log10(ab)   -1.57118    0.45019  -3.490 0.000483 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 116.652  on 99  degrees of freedom
Residual deviance:  98.017  on 97  degrees of freedom
AIC: 104.02

Number of Fisher Scoring iterations: 5
```

Rcode:

```
malaria <- read.table("http://www.math.ntnu.no/~jarlet/statmod/malaria.dat")

attach(malaria)

malreg=glm(mal~age+log10(ab),family=binomial("logit"))

summary(malreg)

probmal=1/(  1+exp (-(  2.57234+ (-0.06546*15)+(-1.57118*log10(1000))    ) )      )
```

2. We see that age is non-significant. We fit a reduced model, and inspect the output

```
malreg2=glm(mal~log10(ab),family=binomial("logit"))

> summary(malreg2)

Call:
glm(formula = mal ~ log10(ab), family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9159  -0.7339  -0.4854   0.8813   2.4722

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.1552     0.8401   2.565 0.010305 *
log10(ab)    -1.6399     0.4449  -3.686 0.000228 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 116.652  on 99  degrees of freedom
Residual deviance:  98.968  on 98  degrees of freedom
AIC: 102.97

Number of Fisher Scoring iterations: 4
```

3. Probability of malaria (wihtout age):

$$p = \frac{1}{1 + e^{-(2.1552+(-1.6399 \log_{10} \text{ab}))}} \tag{5}$$

Plotting p against antibody level (ab):

R code:

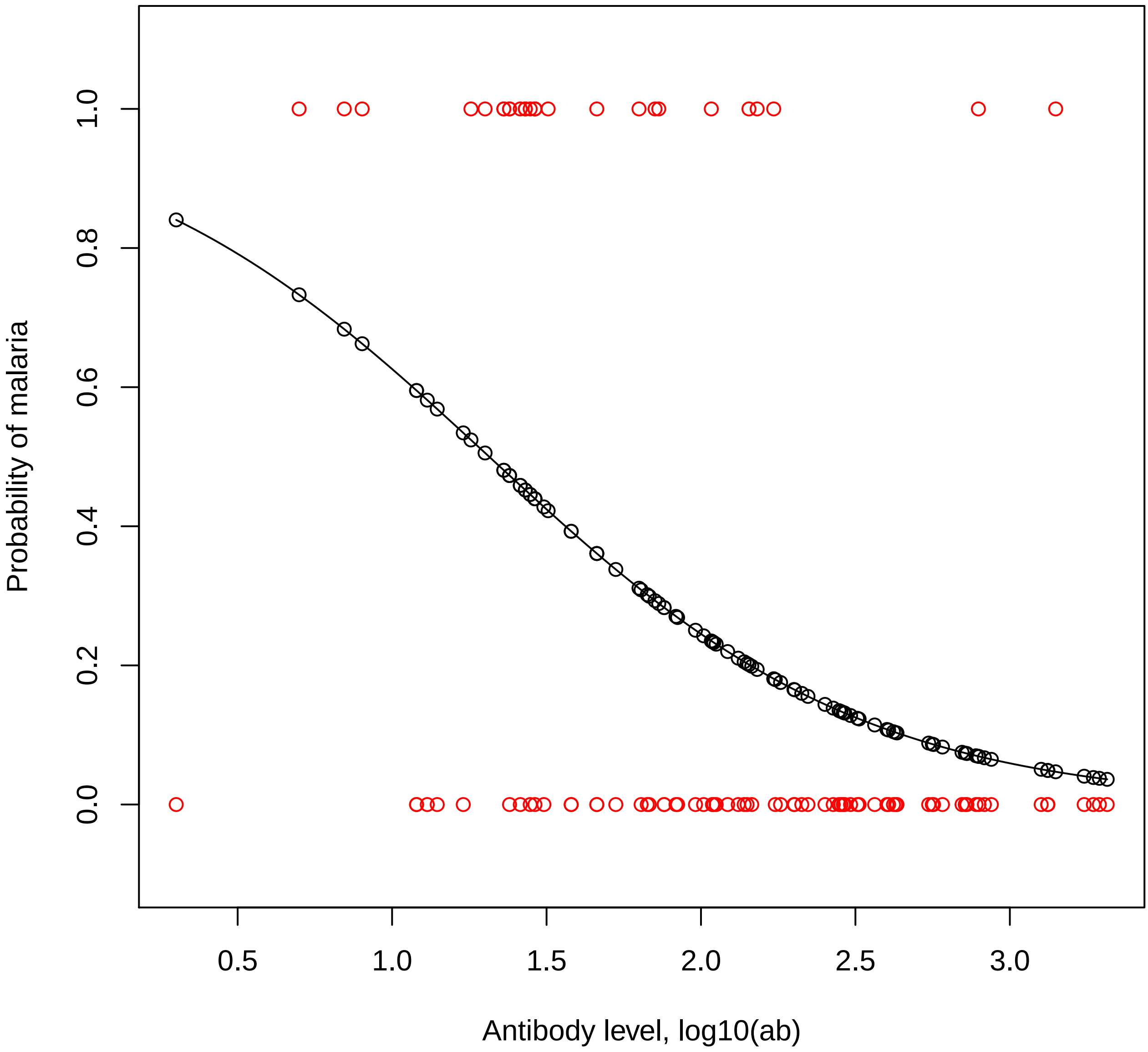

Figure 1: Probablity of malaria against antibody level (on log10 scale) in black and observed data of antibody level againts malaria in red

```
probmal=1/(1+exp(-(2.1552+ (-1.6399*log10(ab)))))

#Plotting the probability againts ab
plot(log10(ab),probmal,ylab="Probability of malaria", xlab="Antibody level, log10(a
#adding a curve
curve(1/(1+exp(-(2.1552+ (-1.6399*x)))),ylab="Probability of malaria", xlab="Antibo
##add observed values of ab
points(x=log10(ab),y=mal,col="RED")
#saving the plot (in the current directory)
dev.copy2pdf(file="plot1oving6.pdf")
```

4. The regression coefficient β for log antibody level represents the the increase in logit(p) (or log odds) for a unit change in $\log_{10}$ antibody level equivalent to a 10-fold increase in antibody level. The odds thus change by an oddsratio equal to exp(β). Based on the estimate of β, the estimate of the oddsratio becomes exp(-1.6399) = 0.1940.

5. Using $confint$ on the fitted model:

```
##CI for the regression coefficients of the fitted model
 10^(confint(malreg2))
Waiting for profiling to be done...
                  2.5 %       97.5 %
(Intercept) 3.797791202 8029.6355942
log10(ab)   0.002567519    0.1483947

##CI for the odds ratio of the fitted model
>  exp(confint(malreg2))
Waiting for profiling to be done...
                 2.5 %     97.5 %
(Intercept) 1.78520113 49.6349805
log10(ab)   0.07498356  0.4366682
```

We see that the confidence interval contains the estimate.