

# Homework4

Anshuman Narayan

11/4/2018

## Q1

We assume that the data is generated from a Binomial logistic regression model where the observed sample proportions of damage are given as  $y_1, \dots, y_5$  are realizations of the independent random variables  $Y_1, \dots, Y_5$  which are defined as the following:  $n_i Y_i \sim \text{Binom}(n_i, \pi(x_{i1}))$  where  $n_1 = n_2 = \dots = n_s = 30$  and  $\pi(x_{i1}) = \exp(\beta_0 + \beta_1 * x_{i1}) / (1 + \exp(\beta_0 + \beta_1 * x_{i1}))$  where  $i=1, \dots, 5$ . where  $\beta_0, \beta_1 \in \mathbb{R}$  are unknown and  $x_{i1}$  is the  $i$ th concentration of the pesticide.  $\beta_0, \beta_1$  are estimated using maximum likelihood.

```
pest=read.table("bliss.txt")
pest
```

```
##   dead alive conc
## 1    2    28    0
## 2    8    22    1
## 3   15    15    2
## 4   23     7    3
## 5   27     3    4
```

```
mod=glm(cbind(dead,alive)~ conc, data=pest , family = binomial)
summary(mod)
```

```
##
## Call:
## glm(formula = cbind(dead, alive) ~ conc, family = binomial, data = pest)
##
## Deviance Residuals:
##      1       2       3       4       5
## -0.4510  0.3597  0.0000  0.0643 -0.2045
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.3238     0.4179  -5.561 2.69e-08 ***
## conc           1.1619     0.1814   6.405 1.51e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 64.76327  on 4  degrees of freedom
## Residual deviance:  0.37875  on 3  degrees of freedom
## AIC: 20.854
##
## Number of Fisher Scoring iterations: 4
```

## Q2

Let  $\hat{\pi}(x)$  be the estimate of  $\pi(x)$ , which is the Binomial logistic regression model estimate of the probability of the death of an insect when the insecticide concentration level is  $x$ .  $\hat{\pi}(x)$  is given by  $\hat{\pi}(x) = \exp(\beta_0 +$

$\beta_1 * x) / (1 + \exp(\beta_0 + \beta_1 * x))$  From our model we have,  $\hat{\beta}_0 = -2.3238$   $\hat{\beta}_1 = 1.1619$  Thus,  $\hat{\pi}(x)$  becomes  $\exp(2.3238 + 1.1619 * x) / (1 + \exp(2.3238 + 1.1619 * x))$

### Q3

The probability of death of an insect exposed to the insecticide at a concentration level of 2 units is given as :

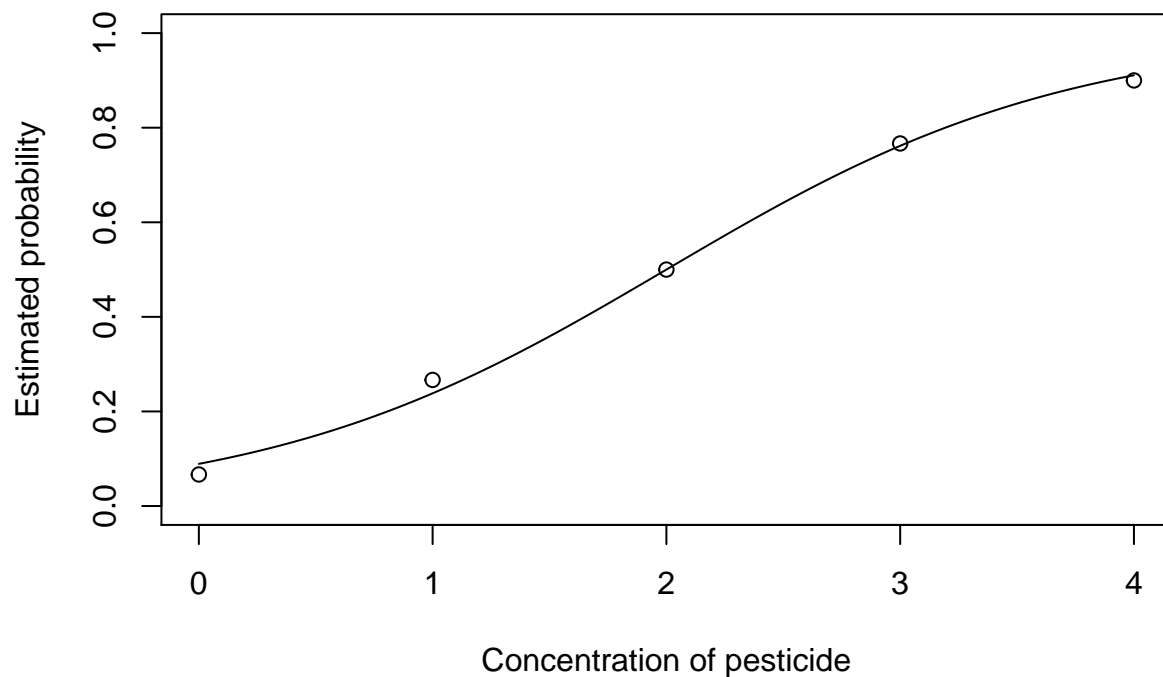
```
ilogit=function(u) return (exp(u)/(1+exp(u)))
cat("\nThe probabily of death of an insect exposed to this insecticide at concentration level of 2 units is\n")

##
## The probabily of death of an insect exposed to this insecticide at concentration level of 2 units is
```

### Q4

The value of  $\hat{\beta}_1$  can be interpreted in terms of log-odds or odds as the following: 1. A unit increase in insecticide concentration level increases the estimated log-odds of insect death by 1.1619 2. A unit increase in insecticide concentration level multiplies the estimated odds of insect death by  $\exp(1.1619)=3.196$ . ###Q5

```
x=seq(from=0,to=4,by=0.01)
pi.hat.x=ilogit(-2.3238+1.1619*x)
plot(dead/30 ~ conc, data=pest,xlim=c(0,4),ylim=c(0,1),xlab="Concentration of pesticide",ylab="Estimated probability")
lines(x,pi.hat.x)
```



### Q6

$H_0 = \exp(\beta_0 + \beta_1 * x_{i1}) / (1 + \exp(\beta_0 + \beta_1 * x_{i1}))$   $H_a$  = The saturated model, without the restriction of  $H_0$  holds

```
#the estimated expected counts when H0 is true
n.list=pest$dead+pest$alive
```

```
expected=cbind(n.list*fitted(mod),n.list*(1-fitted(mod)))
expected
```

```
##           [,1]      [,2]
## 1  2.675153 27.324847
## 2  7.149694 22.850306
## 3 15.000000 15.000000
## 4 22.850306  7.149694
## 5 27.324847  2.675153
```

```
observed=cbind(pest$dead,pest$alive)
```

```
cat("The deviance between expected and observed results is",2*sum(observed * log(observed/expected), na
```

```
## The deviance between expected and observed results is 0.3787483
```

Assuming that  $H_0$  is true, the deviance is a realization of a random variable that is roughly approximated by the Chi-squared distribution with  $n-p-1 = 5-1-1=3$  degrees of freedom.

```
cat("The p-value for this test is ",(1-pchisq(0.3787483,3)))
```

```
## The p-value for this test is  0.9445967
```

This value leads us to reject  $H_0$  and conclude that the current 2 parameter model fits the data well.

## Q7

$H_0 : M_0$  is correct i.e  $\beta_1 = 0$   $H_a : M_f$  without  $M_0$  is correct,  $\beta_1 \neq 0$

```
cat("The deviance for the full model is ",deviance(mod))
```

```
## The deviance for the full model is  0.3787483
```

```
null.mod = glm(cbind(dead, alive) ~ 1,data=pest,family=binomial)
```

```
cat("\nThe deviance for the null model is",deviance(null.mod))
```

```
##
```

```
## The deviance for the null model is 64.76327
```

```
cat("\nThe g-squared random variable for our Likelihood ratio test is",deviance(null.mod)-deviance(mod),
```

```
##
```

```
## The g-squared random variable for our Likelihood ratio test is 64.38452
```

```
cat("\nThe p-value of this test is",1-pchisq(64.38452,1))
```

```
##
```

```
## The p-value of this test is 9.992007e-16
```

This p-value tells us that we can reject  $H_0$  at any practical significance level. ###Q8

```
confint(mod,level=0.99)
```

```
## Waiting for profiling to be done...
```

```
##           0.5 %    99.5 %
## (Intercept) -3.512236 -1.335376
## conc        0.7345749  1.680015
```

The 99% profile likelihood approximate confidence interval for  $\beta_1$  is ( 0.7345749,1.680015)

Q9

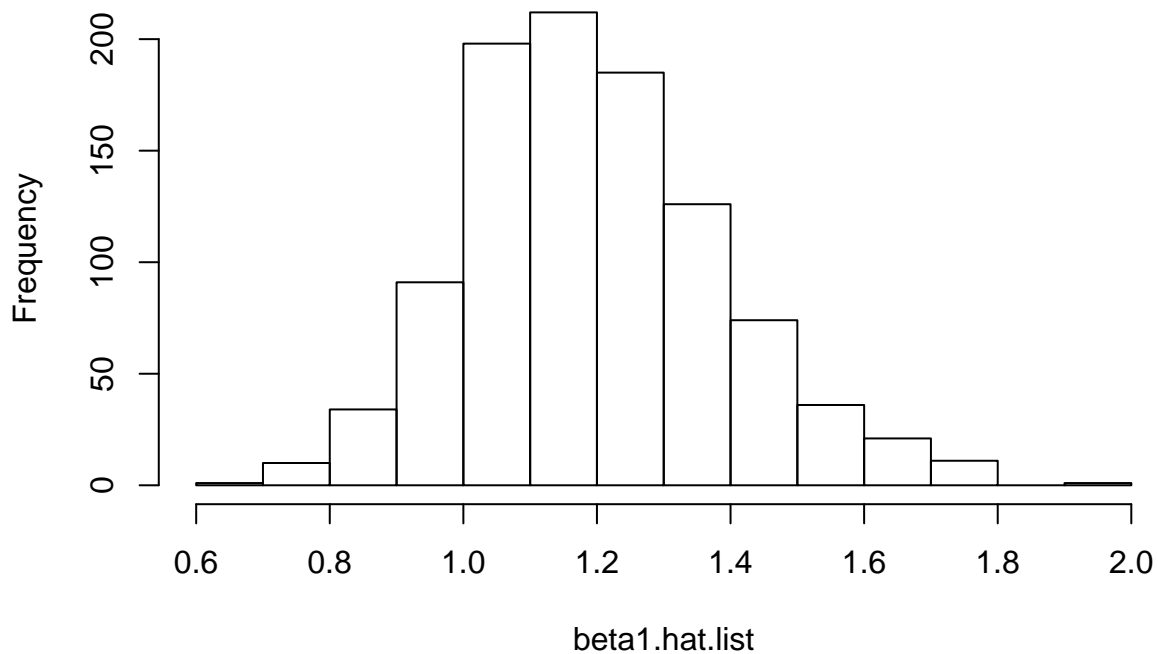
```
dead.sim=rbinom(5,size=30,prob=fitted(mod))
alive.sim=30-dead.sim
conc=0:4
newdat=cbind(dead.sim,alive.sim,conc)
newdat

##      dead.sim alive.sim conc
## [1,]      2      28     0
## [2,]     11      19     1
## [3,]     15      15     2
## [4,]     26       4     3
## [5,]     26       4     4

N=1e3
conc=0:4
beta1.hat.list=NULL
for(r in 1:N)
{
  dead.sim=rbinom(5,size=30,prob=fitted(mod))
  alive.sim=30-dead.sim
  sim.mod=glm(cbind(dead.sim,alive.sim)~conc,family=binomial)
  beta1.hat.list[r]=coef(sim.mod)[2]
}

hist(beta1.hat.list)
```

**Histogram of beta1.hat.list**



```
sd(beta1.hat.list)
```

```
## [1] 0.1934
```

```
summary(mod)
```

```
##
## Call:
## glm(formula = cbind(dead, alive) ~ conc, family = binomial, data = pest)
##
## Deviance Residuals:
##      1      2      3      4      5
## -0.4510  0.3597  0.0000  0.0643 -0.2045
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.3238     0.4179  -5.561 2.69e-08 ***
## conc           1.1619     0.1814   6.405 1.51e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 64.76327  on 4  degrees of freedom
## Residual deviance:  0.37875  on 3  degrees of freedom
## AIC: 20.854
##
## Number of Fisher Scoring iterations: 4
```

The parametric estimate of the standard error of  $\beta_1$  is 0.1943075 while the Wald estimated standard error is 0.1814 which is close to the bootstrap estimate but is lesser.