Assignment 1 - MATH 523 Generalized Linear Models

YUNHEUM DAN SEOL

2018-02-05

1

The inverse Gaussian distribution has density of the form

$$f(y;\mu,\lambda) = (\frac{\lambda}{2\pi y^3})^{1/2}\exp(-\frac{\lambda(y-\mu)^2}{2\mu^2 y})$$

for

$$y > 0$$

and

$$\lambda > 0$$

##(a)A random variable

$$Y \sim f(y;\theta,\lambda)$$

belongs to the exponential dispersion family if

$$f(y) = \exp(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi))$$

for some function

$$a(\cdot), b(\cdot)$$

and

$$c(\cdot)$$

We know if

$$Y \sim inv.Gaussian(y;\mu,\lambda)$$

then the density function of interest would be

$$f(y;\mu,\lambda) = (\frac{\lambda}{2\pi y^3})^{1/2}\exp(-\frac{\lambda(y-\mu)^2}{2\mu^2 y})$$

We use the property

$$x = e^{ln(x)} \quad \forall x \in \mathbb{R}$$

then we obtain

$$(\frac{\lambda}{2\pi y^3})^{1/2}\exp(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}) = \exp(ln(\frac{\lambda}{2\pi y^3}^{1/2}) * \exp(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}) = \exp(-\frac{\lambda(y-\mu)^2}{2\mu^2 y} + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3}))$$

and

$$\exp(-\frac{\lambda(y-\mu)^2}{2\mu^2 y} + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3})) = \exp(-\frac{\lambda(y^2 - 2y\mu + \mu^2)}{2\mu^2 y} + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3}))$$

$$= \exp(\lambda[\frac{-y^2}{2\mu^2 y} + \frac{2y\mu}{2\mu^2 y} + \frac{-\mu^2}{2\mu^2 y}] + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3})) = \exp(\lambda[\frac{-y}{2\mu^2} + \frac{1}{\mu} + \frac{-1}{2y}] + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3}))$$

and then we have

$$\exp(\lambda[\frac{-y}{2\mu^2} + \frac{1}{\mu} + \frac{-1}{2y}] + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3})) = \exp(\frac{-\lambda y}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2y} + \frac{1}{2}ln(\frac{\lambda}{2\pi y^3})) =$$

$$\exp(\frac{-\lambda y}{2\mu^2}+\frac{\lambda}{\mu}+[-\frac{\lambda}{2y}+\frac{1}{2}ln(\frac{\lambda}{2\pi y^3})]) = \exp(\frac{\lambda}{2}[\frac{-y}{\mu^2}+\frac{2}{\mu}]+[-\frac{1}{2y}+\frac{1}{2}ln(\frac{\lambda}{2\pi y^3})]) =$$
$$\exp(\frac{[\frac{-y}{\mu^2}+\frac{2}{\mu}]}{\frac{2}{\lambda}}+[-\frac{1}{2y}+\frac{1}{2}ln(\frac{\lambda}{2\pi y^3})]) = \exp(\frac{[y*(-\frac{1}{\mu^2})+\frac{2}{\mu}]}{\frac{2}{\lambda}}+[-\frac{1}{2y}+\frac{1}{2}ln(\frac{\lambda}{2\pi y^3})])$$

then we have the canonical parameter λ and the dispersion parameter ϕ as below:

$$\theta = \frac{-1}{\mu^2}, \phi = \lambda$$

and we have obtained

$$a(\phi) = \frac{2}{\lambda}, b(\theta) = -2\sqrt{-\theta},\ c(y,\phi) = [-\frac{1}{2y}+\frac{1}{2}ln(\frac{\lambda}{2\pi y^3})]$$

Since

$$\phi$$

is unknown inverse Gaussian distribution belongs to the exponential dispersion family.

(b) from lectures we know

$$\mathbb{E}[\mathrm{Y}] = b'(\theta) = \frac{d}{d\theta}[b(\theta)]$$

and

$$\mathrm{Var}[\mathrm{Y}] = a(\phi)b''(\theta) = a(\phi)v(\mu)$$

; for Inverse Gaussian we have obtained from part (a) that

$$a(\phi) = \frac{2}{\lambda}$$

,

$$b(\theta) = -\sqrt{-2\theta}$$

, and

$$\theta = \frac{-1}{\mu^2}$$

so

$$\frac{d}{d\theta}[b(\theta)] = \frac{d}{d\theta}[-2\sqrt{-\theta}] = -2\frac{d}{d\theta}[\sqrt{-\theta}] = -2(\frac{1}{\sqrt{-\theta}})*(\frac{1}{2})*(-1) =$$
$$\frac{1}{\sqrt{-\theta}} = (-\theta)^{-1/2} = (-(\frac{-1}{\mu^2}))^{-1/2} = (\frac{1}{\mu^2})^{-1/2} = (\mu^{-2})^{-1/2} = \mu$$

and

$$b''(\theta) = \frac{d}{d\theta}[b'(\theta)] = \frac{d}{d\theta}[(-\theta)^{-1/2}] = \frac{-1}{2}*(-\theta)^{-3/2}*(-1) = \frac{1}{2}*(-\theta)^{-3/2} =$$
$$\frac{1}{2}*(-\theta)^{-3/2} = \frac{1}{2}*(\mu^{-2})^{-3/2} = \frac{\mu^3}{2}$$

so we have

$$\mathbb{E}[Y] = \mu$$

,

$$v(\mu) = \frac{\mu^3}{2}$$

,

$$\mathrm{Var}[Y] = a(\phi)*v\mu = \frac{2}{\lambda}*\frac{\mu^3}{2} = \frac{\mu^3}{\lambda}$$

.

(c) We have our

$$\theta = -\frac{1}{\mu^2}$$

. Since we can omit the negative sign by multiplying

$$\beta_j$$

's by negative sign, we can take the canonical link function

as

$$\theta = g(\mu_i) = \frac{1}{\mu_i^2}$$

(Square-reciprocal) So our modeling assumption for inverse Gaussian distribtuion GLM using the canonical link would be

$$g(\mu_i) = \frac{1}{\mu_i^2} = \mathbb{X}\beta$$

.

This link function might not be as sensible as we think because $\frac{1}{x^2}$ is always positive if defined if $\mu \in \mathrm{R}^+$ but our estimators β_j's can have negative components.

Therefore,

$$g(\mu) = ln(\mu)$$

might be a much more reasonable choice.

(d) from class, we know that the score equations (likelihood equations) for any GLM using its canonical link would be of the form

$$\frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n}[\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] * \frac{1}{g'(\mu_i)} * x_{ij} = 0$$

$$\implies \frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n}[\frac{y_i - \mu_i}{a(\phi)} x_{ij}] = 0$$

Since $a(\phi) = \frac{2}{\lambda}$ for Y $\sim inv.Gaussian(\mu, \lambda)$, our score equations would be of the form

$$\implies \frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n}\frac{2}{\lambda}[y_i - \mu_i]x_{ij} = 0$$

##(e) (a GLM with inverse Gaussian responses and the canonical link assumed, since the generalization will be made in A2) We get

$$\frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n}\frac{2}{\lambda}[y_i - \mu_i]x_{ij} = 0$$

$$\implies \frac{\partial l(\theta)}{\partial \beta_0} = \sum_{i=1}^{n}\frac{2}{\lambda}[y_i - \mu_i]$$

and

$$\frac{\partial l(\theta)}{\partial \beta_1} = \sum_{i=1}^{n}\frac{2}{\lambda}[y_i - \mu_i]x_i = \sum_{i=1}^{n_A}\frac{2}{\lambda}[y_i - \mu_i] + \sum_{i=n_A+1}^{n_A+n_B}\frac{2}{\lambda}[y_i - \mu_i] * 0 = 0$$

we have

$$g(\mu_i) = \beta_0 + \beta_1 x_i \implies \mu_i = g^{-1}(\beta_0 + \beta_1 x_i)$$

with our canonical link

$$g(x) = \frac{1}{x^2} \implies g^{-1}(x) = \frac{1}{\sqrt{x}}$$

so we will have

$$\mu_i = \frac{1}{\sqrt{\beta_0 + \beta_1 x_i}}$$

. For group A, $x_i = 1,\ i = 1, ..., n_A$ so we would have

$$\mu_i = \frac{1}{\sqrt{\beta_0 + \beta_1}}$$

For group B, $x_i = 0,\ i = n_A + 1, ..., n_A + n_B$, so we would have

$$\mu_i = \frac{1}{\sqrt{\beta_0}}$$

We know that with canonical link function

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n_A} y_i + \sum_{j=n_A+1}^{n_A+n_B} y_j = \sum_{i=1}^{n_A} \mu_i + \sum_{j=n_A+1}^{n_A+n_B} \mu_j =$$

$$\frac{n_A}{\sqrt{\beta_0 + \beta_1}} + \frac{n_B}{\sqrt{\beta_0}}$$

then

our fitted group means would be

$$\widehat{\mu_{n_A}} = \frac{1}{n_A} \sum_{i=1}^{n_A} y_i = \frac{1}{n_A} \frac{n_A}{\sqrt{\beta_0 + \beta_1}} = \frac{1}{\sqrt{\beta_0 + \beta_1}}$$

$$\widehat{\mu_{n_A}} = \frac{1}{n_A} \sum_{i=1}^{n_A} y_i = \frac{1}{n_A} \frac{n_A}{\sqrt{\beta_0 + \beta_1}} = \frac{1}{\sqrt{\beta_0 + \beta_1}}$$

$$\widehat{\mu_{n_B}} = \frac{1}{n_B} \sum_{j=n_A+1}^{n_A+n_B} y_j = \frac{1}{n_B} \frac{n_B}{\sqrt{\beta_0}} = \frac{1}{\sqrt{\beta_0}}$$

.

2.

$$\frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] * \frac{1}{g'(\mu_i)} * x_{ij} = 0$$

$$\implies \frac{\partial l(\theta)}{\partial \beta_0} = \sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] * \frac{1}{g'(\mu_i)} = 0$$

$$\implies \frac{\partial l(\theta)}{\partial \beta_1} = \sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] * \frac{1}{g'(\mu_i)} * x_i = 0 = \sum_{i=1}^{n_A} \frac{y_i - \mu_i}{\mathrm{Var[y_i]}} * \frac{1}{g'(\mu_i)}$$

We know that $\frac{1}{g'(\mu_i)}$, $\mathrm{Var[Y_i]}$ are nonzero, we know

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n_A} y_i + \sum_{j=n_A+1}^{n_A+n_B} y_j = \sum_{i=1}^{n_A} \mu_i + \sum_{j=n_A+1}^{n_A+n_B} \mu_j$$

and we have for $i = 1, ..., n_A$

$$\mu_i = g^{-1}(\beta_0 + \beta_1)$$

while for $j = n_A + 1, ..., n_A + n_B$ we have

$$\mu_j = g^{-1}(\beta_0)$$

It follows that

$$\sum_{i=1}^{n_A} y_i = \sum_{i=1}^{n_A} \mu_i = n_A g^{-1}(\beta_0 + \beta_1)$$

and

$$\sum_{j=n_A+1}^{n_A+n_B} y_j = \sum_{j=n_A+1}^{n_A+n_B} \mu_j = n_B g^{-1}(\beta_0)$$

thus our fitted group means should be

$$\frac{1}{n_A}\sum_{i=1}^{n_A} y_i = \frac{n_A}{n_A} g^{-1}(\beta_0 + \beta_1) = g^{-1}(\beta_0 + \beta_1)$$

and

$$\frac{1}{n_B}\sum_{j=n_A+1}^{n_A+n_B} y_j = \frac{n_B}{n_B} g^{-1}(\beta_0) = g^{-1}(\beta_0)$$

.

3.

$$\frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] * \frac{1}{g'(\mu_i)} * x_{ij} = 0$$

and we have

$$\frac{1}{g'(\mu_i)} = \frac{\partial \mu_i}{\partial \eta_i},\ x_{ij} = \frac{\partial \eta_i}{\partial \beta_j}$$

$$\sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[Y_i]}}] * \frac{1}{g'(\mu_i)} x_{ij} = \sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} =$$

$$\sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] \frac{\partial \mu_i}{\partial \beta_j} = 0$$

We can show that this comes from the generalized least squares linear model, if we set the link to be the identity link

$$\mu_i = X_i \beta$$

where X_i would be the ith row of the design matrix $\mathbb{X}$, and β the vector parameter to estimate. We set partials to zero in order to minimize our expression

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mathrm{Var}[y_i]} = 0$$

. But

$$0 = \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mathrm{Var}[y_i]} = -2 \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\mathrm{Var}[y_i]} * x_{ij}$$

$\forall j = 1, ..., p$ but we have

$$x_{ij} = \frac{\partial \mu_i}{\partial \beta_j}$$

for our linear model!

So we recover the expression

$$\sum_{i=1}^{n} [\frac{y_i - \mu_i}{\mathrm{Var[y_i]}}] \frac{\partial \mu_i}{\partial \beta_j} = 0$$

(b)

For any GLM with the canonical link with independent $a(\phi)$, we have

$$\frac{\partial l(\theta)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{[y_i - \mu_i]}{a(\phi)} x_{ij} = 0$$

We obtain

$$a(\phi) \sum_{i=1}^{n} \frac{[y_i - \mu_i]}{a(\phi)} x_{ij} = 0 * a(\phi) = 0$$

but

$$a(\phi) \sum_{i=1}^{n} \frac{[y_i - \mu_i]}{a(\phi)} x_{ij} = \sum_{i=1}^{n} [y_i - \mu_i] x_{ij} = 0$$

$\forall j = 1, ..., p$ then it follows that

$$\implies \mathbb{X}(\vec{y} - \vec{\mu}) = \vec{0}$$

So we obtain why the residual vector should be orthogonal to the $colsp(\mathbb{X})$.

4.

(a)

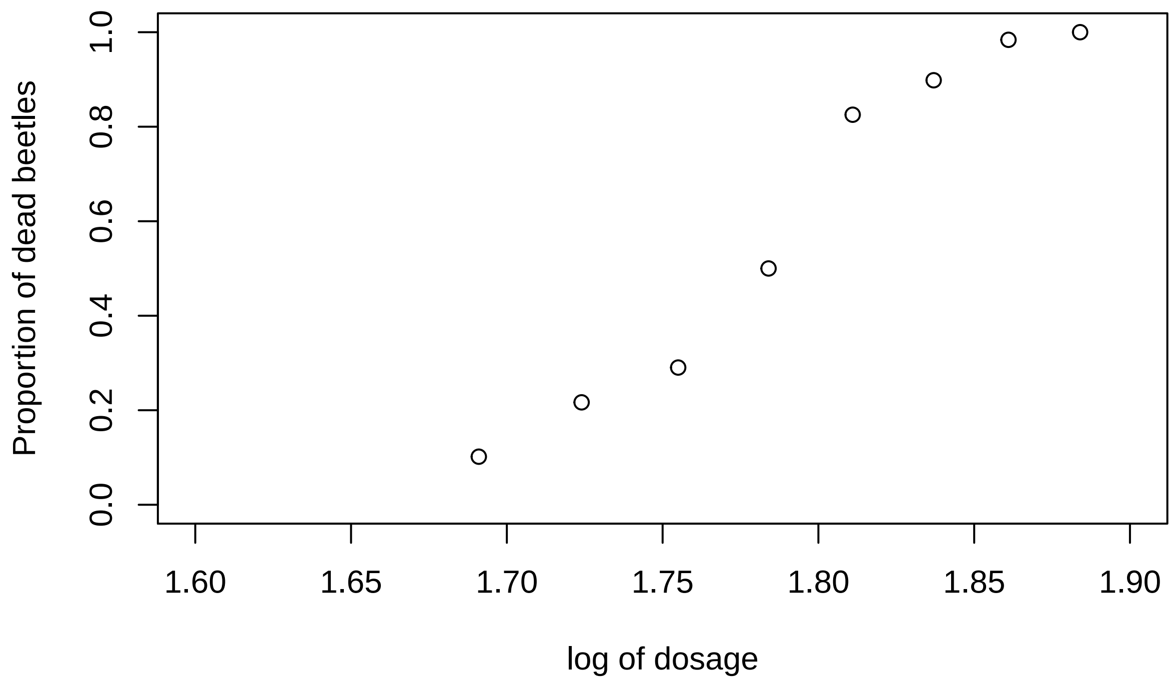
```r
beetles <- read.table("Beetles2.dat", header=TRUE)
attach(beetles)
head(beetles)
```

```
##   logdose  n dead
## 1   1.691 59    6
## 2   1.724 60   13
## 3   1.755 62   18
## 4   1.784 56   28
## 5   1.811 63   52
## 6   1.837 59   53
```

```r
summary(beetles)
```

```
##     logdose            n              dead      
##  Min.   :1.691   Min.   :56.00   Min.   : 6.00  
##  1st Qu.:1.747   1st Qu.:59.00   1st Qu.:16.75  
##  Median :1.798   Median :60.00   Median :40.00  
##  Mean   :1.793   Mean   :60.12   Mean   :36.38  
##  3rd Qu.:1.843   3rd Qu.:62.00   3rd Qu.:54.75  
##  Max.   :1.884   Max.   :63.00   Max.   :61.00  
```

```r
x <- beetles$logdose
y <- beetles$dead/beetles$n
#fit1 <- glm(y~x, family = binomial(link = "logit"))
#summary(fit1)
plot(y~x,xlim=c(1.6,1.9),ylim=c(0,1),xlab="log of dosage",ylab="Proportion of dead beetles")
```




```r
y_a <- cbind(beetles$dead, beetles$n-beetles$dead)

logitfit <- glm(y_a~x, family = binomial(link = "logit"))
summary(logitfit)
```

```
##
## Call:
## glm(formula = y_a ~ x, family = binomial(link = "logit"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5878  -0.4085   0.8442   1.2455   1.5860
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -60.740      5.182  -11.72   <2e-16 ***
## x             34.286      2.913   11.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.202  on 7  degrees of freedom
## Residual deviance:  11.116  on 6  degrees of freedom
## AIC: 41.314
##
```

```
## Number of Fisher Scoring iterations: 4
```

(b)

The fitted value slope coefficient

$$\widehat{\beta_1} = 34.286$$

would the effect of 1 unit increase in log dosage of carbon disulphide increasing the odds of proportion of dead beetles.

(c)

```r
#Next, explore the same GLM model with the probit link

probitfit <-  glm(y_a~x,family=binomial(link=probit))
summary(probitfit)
```

```
##
## Call:
## glm(formula = y_a ~ x, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5627  -0.4848   0.7647   1.0530   1.3149
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -34.956      2.649  -13.20   <2e-16 ***
## x             19.741      1.488   13.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.202  on 7  degrees of freedom
## Residual deviance:   9.987  on 6  degrees of freedom
## AIC: 40.185
##
## Number of Fisher Scoring iterations: 4
```

```r
# Finally, explore the complementary log-log link

loglogfit <- glm(y_a~x,family=binomial(link="cloglog"))

summary(loglogfit)
```

```
##
## Call:
## glm(formula = y_a ~ x, family = binomial(link = "cloglog"))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.80002  -0.56588   0.01475   0.38096   1.31591
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -39.522      3.236  -12.21   <2e-16 ***
## x             22.015      1.797   12.25   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.2024  on 7  degrees of freedom
## Residual deviance:   3.5143  on 6  degrees of freedom
## AIC: 33.712
##
## Number of Fisher Scoring iterations: 4
```

(d)

```r
library(faraway)
logitfit <- glm(y_a~x, family = binomial(link = "logit"))
summary(logitfit)
```

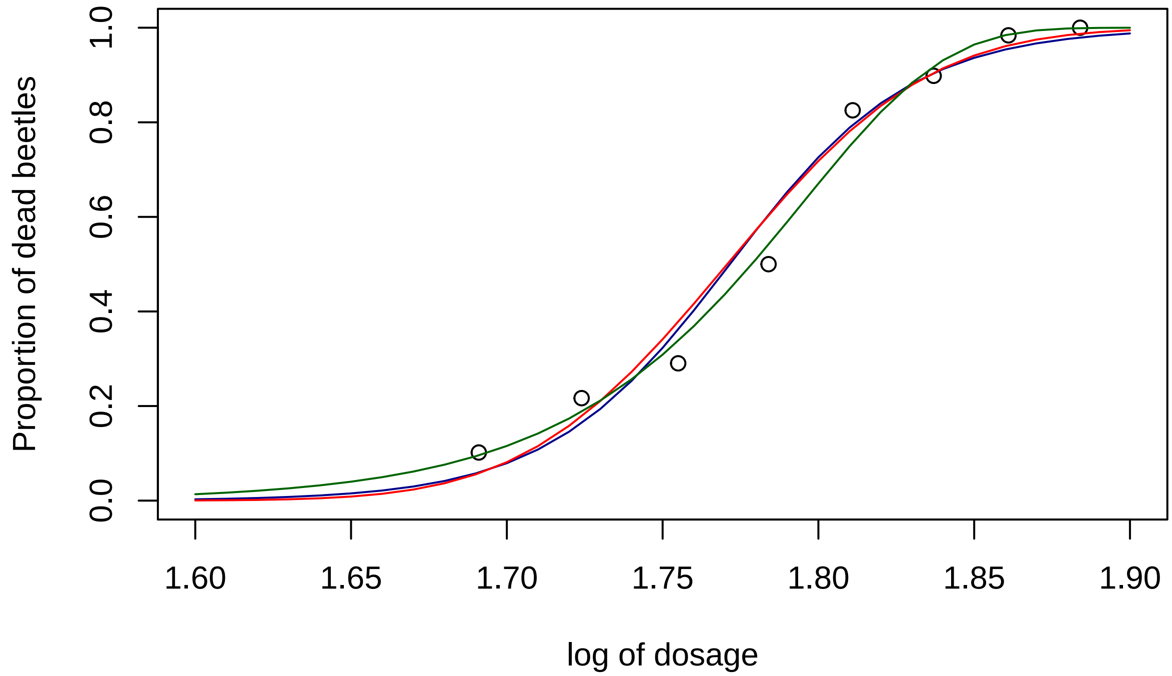
```
##
## Call:
## glm(formula = y_a ~ x, family = binomial(link = "logit"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5878  -0.4085   0.8442   1.2455   1.5860
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -60.740      5.182  -11.72   <2e-16 ***
## x             34.286      2.913   11.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.202  on 7  degrees of freedom
## Residual deviance:  11.116  on 6  degrees of freedom
## AIC: 41.314
##
## Number of Fisher Scoring iterations: 4
```

```r
probitfit <-  glm(y_a~x,family=binomial(link="probit"))
loglogfit <- glm(y_a~x,family=binomial(link="cloglog"))
plot(y~x,xlim=c(1.6,1.9),ylim=c(0,1),xlab="log of dosage",ylab="Proportion of dead beetles")
 x1 <- seq(from=1.6,to=1.9,by=0.01)
lines(x1,ilogit(coef(logitfit)[1]+coef(logitfit)[2]*x1),col="darkblue")
```

```r
lines(x1,pnorm(coef(probitfit)[1]+coef(probitfit)[2]*x1),col="red")



lines(x1,1-exp(-exp(coef(loglogfit)[1]+coef(loglogfit)[2]*x1)),col="darkgreen")
```



##(e)

```r
v1 <- rep(0, 8)
v2 <- rep(0,8)
v3 <- rep(0,8)

for(i in 1:8){
v1[i] <- ilogit(coef(logitfit)[1]+coef(logitfit)[2]*logdose[i])

v2[i] <- pnorm(coef(probitfit)[1]+coef(probitfit)[2]*logdose[i])

v3[i] <- 1-exp(-exp(coef(loglogfit)[1]+coef(loglogfit)[2]*logdose[i]))
}
#fitted values for logit link
v1
```

```
## [1] 0.05937747 0.16366723 0.36162283 0.60490961 0.79440490 0.90405532
## [7] 0.95546748 0.97925643
```

```r
#fitted values for probit link
v2
```

```
## [1] 0.0577367 0.1781060 0.3780390 0.6032833 0.7866532 0.9045852 0.9626183
## [8] 0.9873227
```

```r
#fitted values for complementary log-log link
v3
```

```
## [1] 0.09582195 0.18802653 0.33777217 0.54177644 0.75683967 0.91843509
```

```
## [7] 0.98575181 0.99913561
```

(f)

```r
# model selection with AIC

logitfit$aic
```

```
## [1] 41.31361
```

```r
probitfit$aic
```

```
## [1] 40.18499
```

```r
loglogfit$aic
```

```
## [1] 33.71237
```

Having compared the AIC's for each mode, we can conclude that our model with complementary log-log link seems to be the best fit!