Beetle Example: Logistic Regression with Grouped Data

(Simple) logistic regression is used to model the relationship between a binary response variable and a single numerical predictor. In this example, beetle death is considered an "event" and logdose is the predictor. Note that this data is grouped, meaning that for each of 8 doses we have approximately 60 beetles (either dead or alive) and the data is summarized in only 8 rows (corresponding to 8 doses). The glm function can be used to run logistic regression whether or not the data is grouped, just be careful about the formatting!

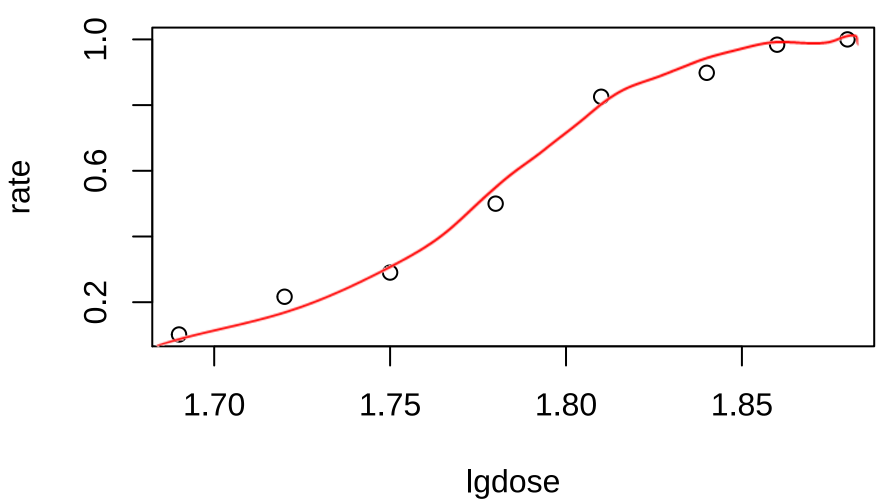
```
library(car)
```

```
## Loading required package: carData
```

```
library(MASS)
Beetle <- read.csv("~/Dropbox/STAT512/Lectures/MultReg5/MR5_Beetles.csv")
Beetle
```

```
##   lgdose nrtest nrdead
## 1   1.69     59      6
## 2   1.72     60     13
## 3   1.75     62     18
## 4   1.78     56     28
## 5   1.81     63     52
## 6   1.84     59     53
## 7   1.86     62     61
## 8   1.88     60     60
```

```
Beetle$rate <- Beetle$nrdead/Beetle$nrtest
plot(rate ~ lgdose, data = Beetle)
```



Logistic Regression model

With different N at each dose level, we provide a 2-column matrix of #successes and failures as the response "variable".

event   not event

```
Model1 <- glm(cbind(nrdead, nrtest-nrdead) ~ lgdose, family = binomial(link = "logit"), data = Beetle)
summary(Model1)
```

```
##
```

```
## Call:
## glm(formula = cbind(nrdead, nrtest - nrdead) ~ lgdose, family = binomial(link = "logit"),
##     data = Beetle)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4242  -0.6084   0.7535   1.1080   1.6837
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -59.282      4.995  -11.87   <2e-16 ***
## lgdose        33.519      2.814   11.91   <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:   9.9971  on 6  degrees of freedom
## AIC: 40.195
##
## Number of Fisher Scoring iterations: 4
```

```
Anova(Model1, type = 3)
```

not needed

```
## Analysis of Deviance Table (Type III tests)
##
## Response: cbind(nrdead, nrtest - nrdead)
##        LR Chisq Df Pr(>Chisq)
## lgdose    274.2  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Odds ratio estimates and CIs
#To compute odds ratio estimates: exponentiate estimates and CI endpoints.
exp(Model1$coef)
```

coefficient for glm() object

```
##  (Intercept)       lgdose
## 1.794846e-26 3.607265e+14
```

```
confint(Model1)
```

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

```
##                 2.5 %    97.5 %
## (Intercept) -69.61409 -49.97808
## lgdose       28.27950  39.34206
```

```
exp(confint(Model1))
```

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

```
##                    2.5 %       97.5 %
## (Intercept) 5.847726e-31 1.971502e-22
## lgdose      1.912621e+12 1.219095e+17
```

do this for reporting

Examine Fit

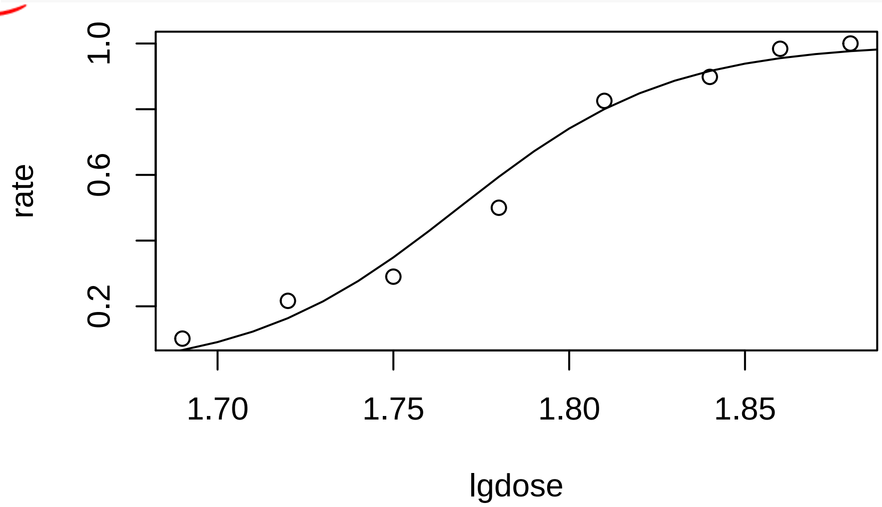
Here we calculate (McFadden's) psuedo R2 and graph the data with the fitted logistic curve overlaid. For the predict function, note that the type = "response" option returns the response on the proportion (or probability) scale instead of the default logit scale.

```
#Calculate McFadden's Pseudo R2 "by hand"
NullModel <- glm(cbind(nrdead, nrtest-nrdead) ~ 1, family = binomial(link = "logit"), data = Beetle)
1-logLik(Model1)/logLik(NullModel)
```

```
## 'log Lik.' 0.883392 (df=2)
```

this is really good dosage w/ small steps

```
#Plot the fitted curve
plot(rate ~ lgdose, data = Beetle)
lgdosenew <- seq(1.66, 1.9, 0.01)
phat <- predict(Model1, list(lgdose = lgdosenew), type = "response")
lines(phat ~ lgdosenew)
```



Estimate LD(p)

install

dose.p() from MASS package computes LD's for various probs. cf=1:2 tells it that coef[1] is the intercept and coef[2] is the slope.

```
probs <- seq(0.1, 0.9, 0.05)
ld <- dose.p(Model1, cf = 1:2, p = probs)
ld
```

seq of probs

for coeffs

```
##              Dose          SE
## p = 0.10: 1.703059 0.007155075
## p = 0.15: 1.716861 0.006217407
## p = 0.20: 1.727252 0.005566667
## p = 0.25: 1.735834 0.005079572
## p = 0.30: 1.743332 0.004703378
## p = 0.35: 1.750142 0.004412019
## p = 0.40: 1.756514 0.004192013
## p = 0.45: 1.762623 0.004036757
## p = 0.50: 1.768610 0.003943864
## p = 0.55: 1.774597 0.003913824
```

$LD_{50} = 1.77$

lethal log dose 50

```
## p = 0.60: 1.780707 0.003949443
## p = 0.65: 1.787078 0.004055959
## p = 0.70: 1.793888 0.004242038
## p = 0.75: 1.801386 0.004522291
## p = 0.80: 1.809968 0.004922971
## p = 0.85: 1.820360 0.005496139
## p = 0.90: 1.834161 0.006363428
```