Beetle Kill Data Example

jdt

3/10/2021

Contents

Background 1

Logistic Models 1
- GLM for Binary Data . . . 1
- Linear Model . . . 2
- Logistic Model . . . 2
- Probit Model . . . 2
- Complementary Log-Log Model . . . 2

R 3

SAS 7
- Code . . . 7
- Output . . . 8

Background

In this document, I use the GLM for binary data to model a dose response result. The example uses the Dobson Beetle Kill Data where increasing levels of a pesticide is studied. Historically, the probit function was used to model data of this type. In this example, the model is compared using the linear, probit, logit, and cloglog links.

Logistic Models

GLM for Binary Data

Let Y be a binary response variable where $\Pr[Y = 1 \mid \mathbf{x}] = \pi(\mathbf{x})$ and $\Pr[Y = 0 \mid \mathbf{x}] = 1 - \pi(\mathbf{x})$ with covariates $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. There are several potential approaches to this modeling problem.

Linear Model

One could use the ordinary least squares approach, called the **Linear Probability model**, given as,

$$\pi(\mathbf{x}) = \alpha + \beta'\mathbf{x}. \tag{1}$$

This model has a structural defect since $\pi(x)$ is not restricted to the interval $[0, 1]$ for all x.

Logistic Model

A better model is the **Logistic Regression Model** given as,

$$y = log\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = (\alpha + \beta'\mathbf{x}), \tag{2}$$

where y is the log odds and $\pi(\mathbf{x})$ is the probability of the event of interest for the covariate **x**. It follows that,

$$\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = exp(\alpha + \beta'\mathbf{x}),$$

and

$$\pi(\mathbf{x}) = \frac{exp(\alpha + \beta'\mathbf{x})}{1 + exp(\alpha + \beta'\mathbf{x})}.$$

Probit Model

Historically, toxicological experiments often measure dosage as the log concentration where the tolerance distribution for the dosage is assumed to be approximately $N(\mu, \sigma^2)$ for unknown μ and σ^2. In which case, the **Probit Model** is given as

$$\pi(x) = \Phi(\alpha + \beta\, x) \tag{3}$$

where Φ is the standard normal cdf, $\alpha = -\mu/\sigma$ and $\beta = 1/\sigma$.

Complementary Log-Log Model

The complementary log-log model provides an alternative model to the logit and probit that is asymmetric about 0.5 where the **Complementary Log-Log Model** is

$$\pi(x) = 1 - \exp[-\exp(\alpha + \beta x)], \tag{4}$$

and

$$\log[-\log(1 - \pi(x))] = \alpha + \beta x.$$

Let x_1 and x_2 denote two values of the covariate, then

$$\log[-\log(1 - \pi(x_2))] - \log[-\log(1 - \pi(x_1))] = \beta(x_2 - x_1),$$

or

$$\frac{\log[1 - \pi(x_2)]}{\log[1 - \pi(x_1)]} = \exp[\beta(x_2 - x_1)].$$

In which case, one has

$$1 - \pi(x_2) = [1 - \pi(x_1)]^{\exp[\beta(x_2 - x_1)]}.$$

R

Needed Packages

```
if(!require(FSA)){install.packages("FSA")}
if(!require(ggplot2)){install.packages("ggplot2")}
if(!require(car)){install.packages("car")}
if(!require(multcompView)){install.packages("multcompView")}
if(!require(lsmeans)){install.packages("lsmeans")}
if(!require(grid)){install.packages("grid")}
if(!require(nlme)){install.packages("nlme")}
if(!require(lme4)){install.packages("lme4")}
if(!require(Rmisc)){install.packages("Rmisc")}
if(!require(rms)){install.packages("rms")}
if(!require(FSA)){install.packages("FSA")}
#if(!require(lmerTest)){install.packages("lmerTest")}
#if(!require(rcompanion)){install.packages("rcompanion")}
```

Read data from SAS input file

```
### --------------------------------------------------------------
### Two-way anova, SAS example Activity by Genotype
### --------------------------------------------------------------

Input = ("
logdose nbeetles nkilled
1.691 59 6
1.724 60 13
1.755 62 18
1.784 56 28
1.811 63 52
1.837 59 53
1.861 62 61
1.884 60 60
")
beetle = read.table(textConnection(Input),header=TRUE)
beetle = data.frame(beetle)

logdose = beetle$logdose
nbeetles = beetle$nbeetles
nkilled = beetle$nkilled
nsurvive=nbeetles-nkilled;

per.killed = nkilled/nbeetles
```

Descriptive Statistics

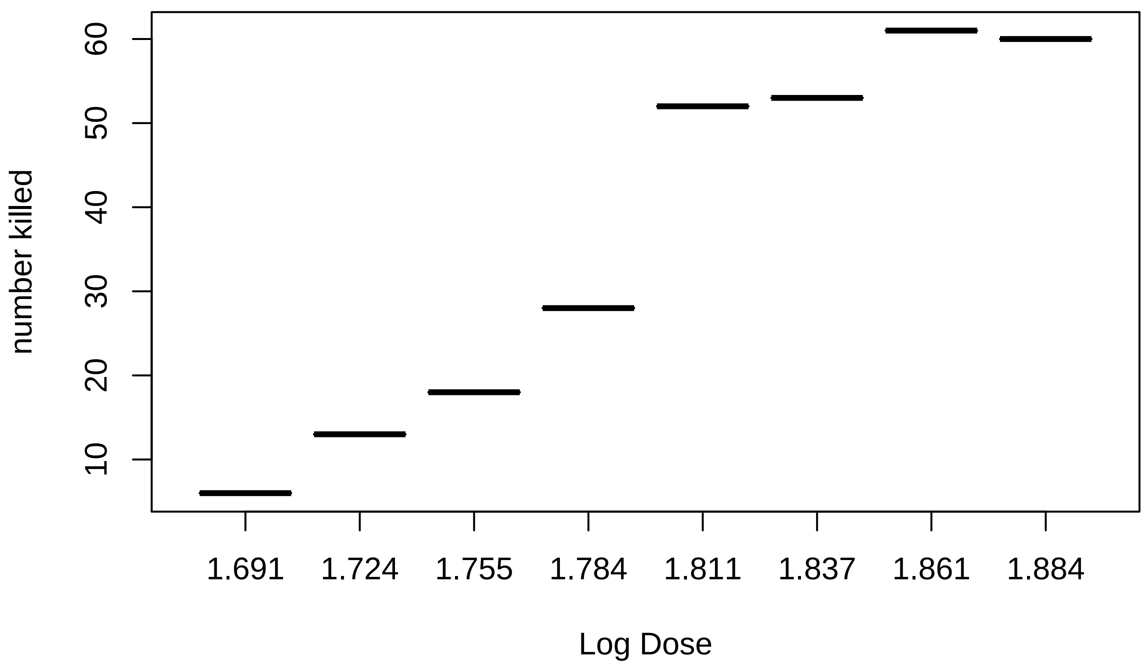
```r
library(Rmisc)

sum = summary(beetle)
sum
```

```
##      logdose          nbeetles        nkilled     
##  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
```

Descriptive Plots

```r
boxplot(nkilled ~ logdose,
        data = beetle,
        xlab = "Log Dose",
        ylab = "number killed")
```



Fit Linear Regression Model

```r
mod1 = lm( per.killed ~ logdose, data=beetle)
mod1
```

```
## 
## Call:
```

```
## lm(formula = per.killed ~ logdose, data = beetle)
##
## Coefficients:
## (Intercept)      logdose
##      -8.948        5.325
```

Fit Logistic Model with PROBIT link

```
mod2 = glm( nkilled/nbeetles ~ logdose,
            family = "binomial"(link="probit"),
            data=beetle)
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
mod2
```

```
##
## Call:  glm(formula = nkilled/nbeetles ~ logdose,
family = binomial(link = "probit"),
##     data = beetle)
##
## Coefficients:
## (Intercept)      logdose
##      -34.82        19.67
##
## Degrees of Freedom: 7 Total (i.e. Null);  6 Residual
## Null Deviance:       4.712
## Residual Deviance: 0.1676    AIC: 8.094
```

```
#plot(mod1)
```

Fit Logistic Model with LOGIT link

```
mod3 = glm( nkilled/nbeetles ~ logdose,
            family = "binomial"(link="logit"),
            data=beetle)
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
mod3
```

```
##
## Call:  glm(formula = nkilled/nbeetles ~ logdose,
family = binomial(link = "logit"),
##     data = beetle)
##
## Coefficients:
## (Intercept)      logdose
##      -60.48        34.14
##
## Degrees of Freedom: 7 Total (i.e. Null);  6 Residual
```

```
## Null Deviance:       4.712
## Residual Deviance: 0.1867    AIC: 8.03
```

```r
#plot(mod1)
```

Fit Logistic Model with Cloglog link

```r
mod4 = glm( nkilled/nbeetles ~ logdose,
            family = "binomial"(link="cloglog"),
            data=beetle)
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```r
mod4
```

```
##
## Call:  glm(formula = nkilled/nbeetles ~ logdose,
family = binomial(link = "cloglog"),
##     data = beetle)
##
## Coefficients:
## (Intercept)      logdose
##      -39.48        21.99
##
## Degrees of Freedom: 7 Total (i.e. Null);  6 Residual
## Null Deviance:       4.712
## Residual Deviance: 0.05737   AIC: 7.759
```

```r
#plot(mod2)
```

List of Predictive Probabilities

```r
#temperature = runif(100,45,85)
#temperature = order(temperature)
p_linear = -8.95 + 5.33*logdose
p_probit = pnorm((-34.82 + 19.82*logdose),0,1)
p_logit = exp(-60.48+34.14*logdose)/(1 + exp(-60.48+34.14*logdose))
p_cll = 1 - exp(-exp(-39.48 + 21.99*logdose))

cbind(logdose, per.killed, p_linear,p_probit, p_logit, p_cll)
```

```
##      logdose per.killed p_linear   p_probit    p_logit      p_cll
## [1,]   1.691  0.1016949  0.06303 0.09605203 0.06012846 0.09585946
## [2,]   1.724  0.2166667  0.23892 0.25774277 0.16484110 0.18795811
## [3,]   1.755  0.2903226  0.40415 0.48568105 0.36255311 0.33745254
## [4,]   1.784  0.5000000  0.55872 0.70501517 0.60486074 0.54110112
## [5,]   1.811  0.8253968  0.70263 0.85859314 0.79372716 0.75596008
## [6,]   1.837  0.8983051  0.84121 0.94400817 0.90336450 0.91777934
## [7,]   1.861  0.9838710  0.96913 0.98053945 0.95497813 0.98552060
## [8,]   1.884  1.0000000  1.09172 0.99414691 0.97895371 0.99910866
```

SAS

Code

```
title1 'Dobson Table 7.2 Beetle mortality';
data beetle; SET SASUSER.dobson_beetle; RUN;
data beetle; set beetle; rate=kill/number;
run;
title2 'GLM Model for Rate';

title2 'Linear';
proc reg data=beetle plots=observedbypredicted;
  model rate = dose;
run;

title2 'PROBIT';
proc logistic data=beetle plots=effect;
model kill/number = dose/link=probit cl ;
run;

title2 'LOGIT';
proc logistic data=beetle plots=effect;
  model kill/number = dose/link=logit expb cl ;
run;

title2 'CLOGLOG';
proc logistic data=beetle plots=effect;
  model kill/number = dose/link=cloglog cl ;
run;
```

Output

Dobson Table 7.2 Beetle mortality
Linear

The REG Procedure

Model: MODEL1

Dependent Variable: rate

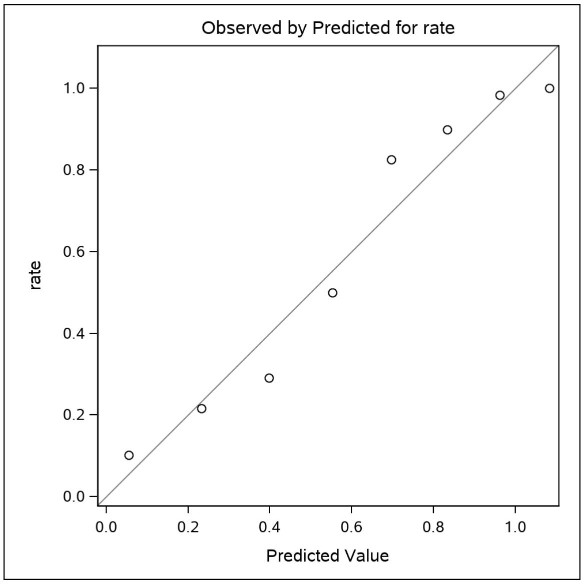
Number of Observations Read	8
Number of Observations Used	8

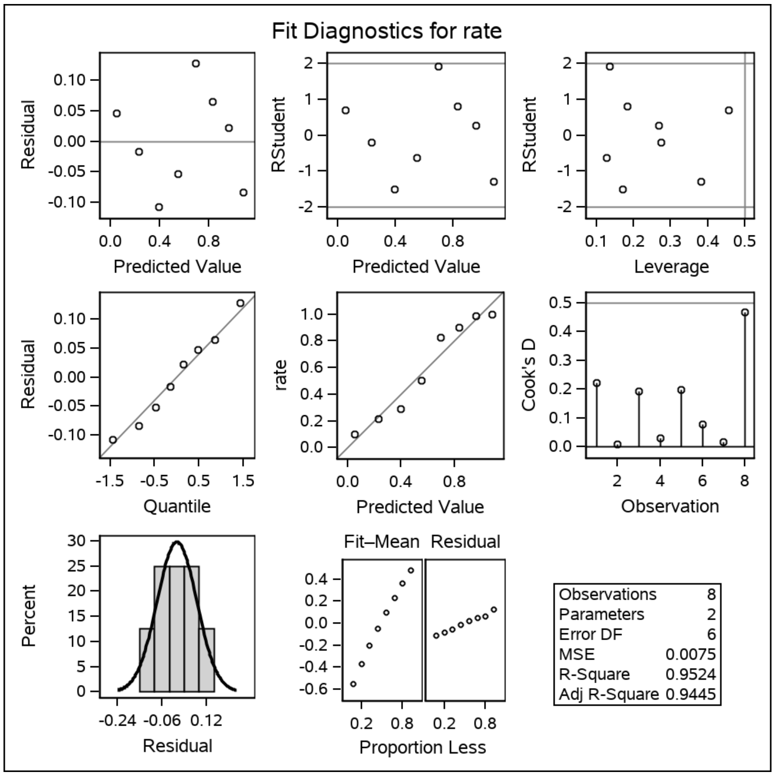
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.90318	0.90318	120.18	<.0001
Error	6	0.04509	0.00752		
Corrected Total	7	0.94827			

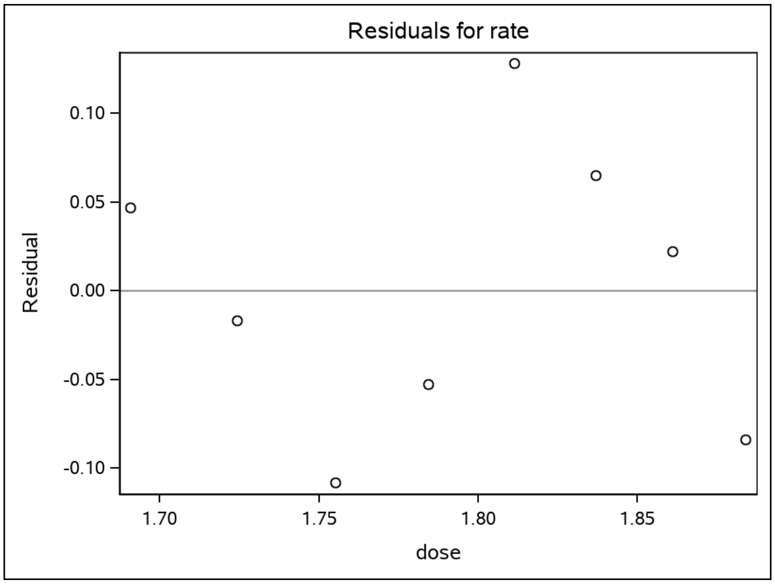
Root MSE	0.08669	R-Square	0.9524
Dependent Mean	0.60203	Adj R-Sq	0.9445
Coeff Var	14.39951		

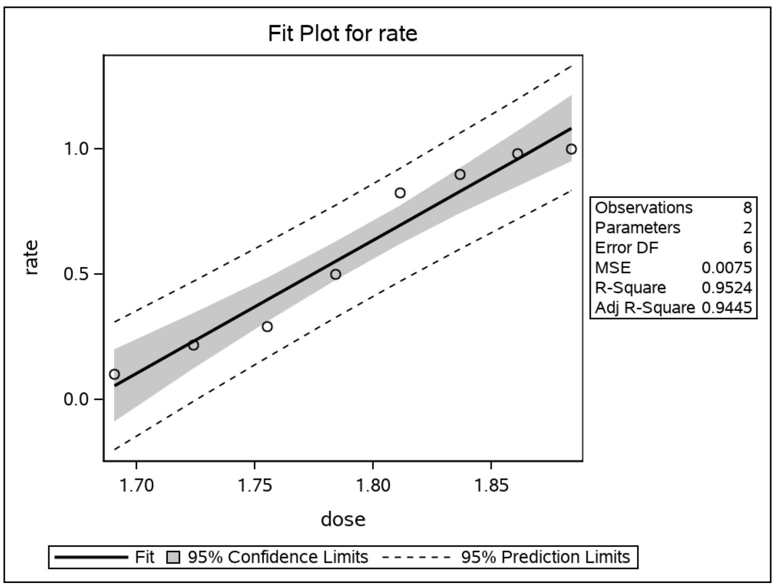
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	−8.94784	0.87166	−10.27	<.0001
dose	1	5.32494	0.48573	10.96	<.0001

Dobson Table 7.2 Beetle mortality
Linear






Residuals for rate
Residual
0.10
0.05
0.00
-0.05
-0.10
1.70
1.75
1.80
1.85
dose


Fit Plot for rate
rate
1.0
0.5
0.0
1.70
1.75
1.80
1.85
dose
Observations 8
Parameters 2
Error DF 6
MSE 0.0075
R-Square 0.9524
Adj R-Square 0.9445
Fit
95% Confidence Limits
95% Prediction Limits

Dobson Table 7.2 Beetle mortality

PROBIT

The LOGISTIC Procedure

Model Information	
Data Set	WORK.BEETLE
Response Variable (Events)	kill
Response Variable (Trials)	number
Model	binary probit
Optimization Technique	Fisher's scoring

Number of Observations Read	8
Number of Observations Used	8
Sum of Frequencies Read	481
Sum of Frequencies Used	481

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	291
2	Nonevent	190

Model Convergence Status
Convergence criterion (GCONV=1E−8) satisfied.

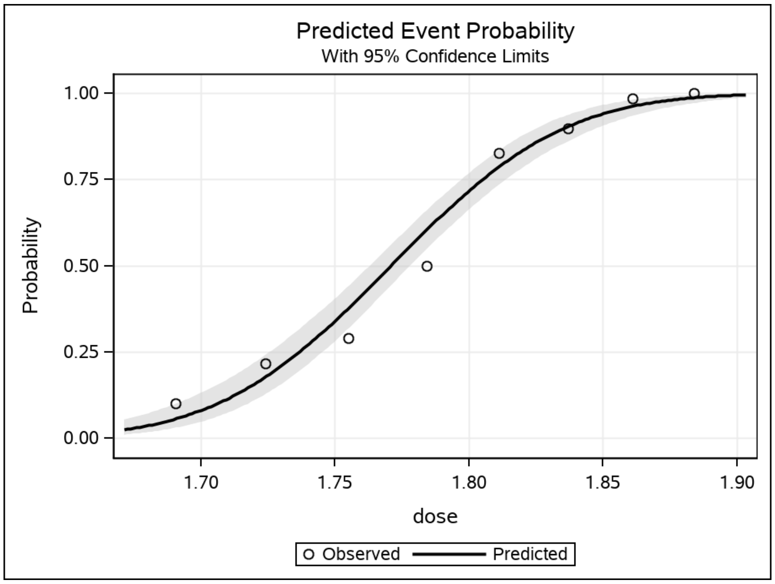
Model Fit Statistics			
Criterion	**Intercept Only**	**Intercept and Covariates**	
		Log Likelihood	**Full Log Likelihood**
AIC	647.441	375.358	40.318
SC	651.617	383.710	48.670
-2 Log L	645.441	371.358	36.318

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	274.0827	1	<.0001
Score	227.5801	1	<.0001
Wald	175.9559	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	−34.9348	2.6479	174.0679	<.0001
dose	1	19.7277	1.4872	175.9559	<.0001

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	87.0	Somers' D	0.802
Percent Discordant	6.8	Gamma	0.856
Percent Tied	6.3	Tau-a	0.384
Pairs	55290	c	0.901

Parameter Estimates and Wald Confidence Intervals			
Parameter	Estimate	95% Confidence Limits	
Intercept	−34.9348	−40.1246	−29.7451
dose	19.7277	16.8128	22.6426



Predicted Event Probability
With 95% Confidence Limits
Probability
1.00
0.75
0.50
0.25
0.00
1.70
1.75
1.80
1.85
1.90
dose
Observed
Predicted

Dobson Table 7.2 Beetle mortality
LOGIT
The LOGISTIC Procedure

Model Information	
Data Set	WORK.BEETLE
Response Variable (Events)	kill
Response Variable (Trials)	number
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	8
Number of Observations Used	8
Sum of Frequencies Read	481
Sum of Frequencies Used	481

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	291
2	Nonevent	190

Model Convergence Status
Convergence criterion (GCONV=1E−8) satisfied.

Model Fit Statistics			
Criterion	**Intercept Only**	**Intercept and Covariates**	
		Log Likelihood	**Full Log Likelihood**
AIC	647.441	376.471	41.430
SC	651.617	384.823	49.782
-2 Log L	645.441	372.471	37.430

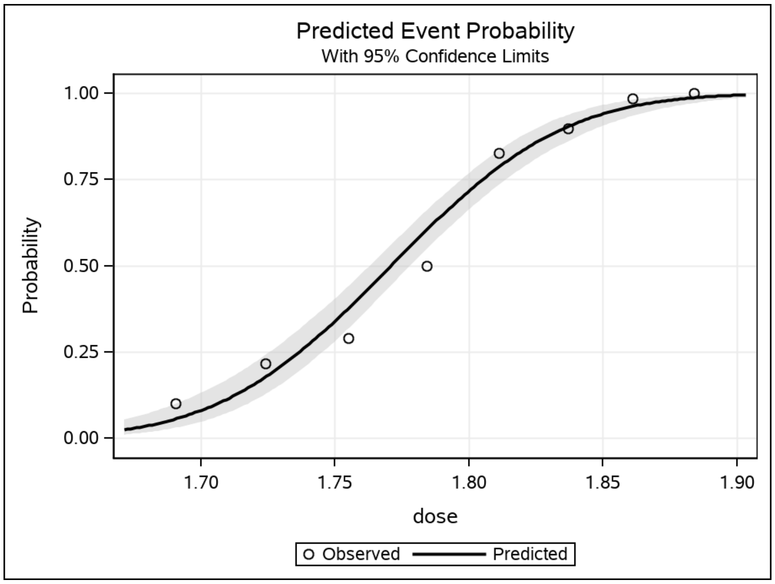
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	272.9702	1	<.0001
Score	227.5801	1	<.0001
Wald	138.4876	1	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	−60.7114	5.1802	137.3557	<.0001	0.000
dose	1	34.2669	2.9118	138.4876	<.0001	7.619E14

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
dose	>999.999	>999.999	>999.999

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	87.0	Somers' D	0.802
Percent Discordant	6.8	Gamma	0.856
Percent Tied	6.3	Tau-a	0.384
Pairs	55290	c	0.901

Parameter Estimates and Wald Confidence Intervals			
Parameter	Estimate	95% Confidence Limits	
Intercept	−60.7114	−70.8644	−50.5584
dose	34.2669	28.5597	39.9740


Predicted Event Probability
With 95% Confidence Limits
Probability
1.00
0.75
0.50
0.25
0.00
1.70
1.75
1.80
1.85
1.90
dose
Observed
Predicted

Dobson Table 7.2 Beetle mortality

CLOGLOG

The LOGISTIC Procedure

Model Information	
Data Set	WORK.BEETLE
Response Variable (Events)	kill
Response Variable (Trials)	number
Model	binary cloglog
Optimization Technique	Fisher's scoring

Number of Observations Read	8
Number of Observations Used	8
Sum of Frequencies Read	481
Sum of Frequencies Used	481

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	291
2	Nonevent	190

Model Convergence Status
Convergence criterion (GCONV=1E−8) satisfied.

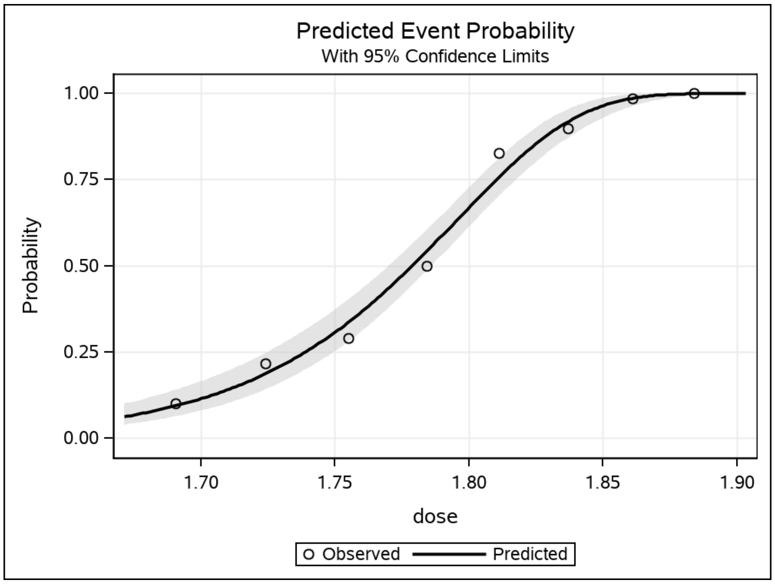
Model Fit Statistics			
Criterion	**Intercept Only**	**Intercept and Covariates**	
		Log Likelihood	**Full Log Likelihood**
AIC	647.441	368.685	33.644
SC	651.617	377.037	41.996
-2 Log L	645.441	364.685	29.644

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	280.7560	1	<.0001
Score	227.5801	1	<.0001
Wald	150.0498	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	−39.5725	3.2403	149.1487	<.0001
dose	1	22.0412	1.7994	150.0498	<.0001

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	87.0	Somers' D	0.802
Percent Discordant	6.8	Gamma	0.856
Percent Tied	6.3	Tau-a	0.384
Pairs	55290	c	0.901

Parameter Estimates and Wald Confidence Intervals			
Parameter	Estimate	95% Confidence Limits	
Intercept	−39.5725	−45.9233	−33.2216
dose	22.0412	18.5146	25.5679


Predicted Event Probability
With 95% Confidence Limits
1.00
0.75
0.50
0.25
0.00
Probability
1.70
1.75
1.80
1.85
1.90
dose
Observed
Predicted