Homework1_601

Vinny Paris

1/21/2018

Please note the code can be downloaded at

1

After reading in the data I created a small function that found the liklihood function for a Binomial random variable with a Pregibon One-Parameter Link. The random variable was the tolerance of the the beetle at a certain level of toxicity. Please see attached hand written sheet for the derivation. After that both a logit and a cloglog model was fit to both series. When it was shown that the cloglog had a smaller liklihood in both samples the decision was made to use it's estimates as the starting value in optim for the linear parameters. The lambda's initial estimate was decided to be half way point between the logit (lambda = 1) and the cloglog (lambda = 0). The first of the two to be presented is the cloglog.

It should be noted that I had to approximate the cloglog value by choosing a sufficently small lambda as it asymptotically tends to the complementary log-log model. Optim was then used to find the MLE's of the parameters. The 1 and 2 after the model/variable name indicate which series it's from.

```
## $par
## [1] -41.5673506  10.0749367   0.1147649
##
## $value
## [1] 11.50023
##
## $counts
## function gradient
##      240       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```
## $par
## [1] -35.2908846   8.4988137  -0.2347339
##
## $value
## [1] 13.46024
##
## $counts
## function gradient
##      212       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```r
#Est vs Logit Series 1
chi1 <- -2*(jj1$value - logitmod1$value)
pchisq(chi1, 1, lower.tail = FALSE)
```

```
## [1] 0.1104696
```

```r
#Est vs Cloglog Series 1
chi1_clog <- -2*(jj1$value - clogmod1$value)
pchisq(chi1_clog, 1, lower.tail = FALSE)
```

```
## [1] 0.7518945
```

```r
#Est vs Cloglog Series 2
chi2_clog <- -2*(jj2$value - clogmod2$value)
pchisq(chi2_clog, 1, lower.tail = FALSE)
```

```
## [1] 0.6174729
```

```r
#Est vs Logit Series 2
chi2 <- -2*(jj2$value - logitmod2$value)
pchisq(chi2, 1, lower.tail = FALSE)
```

```
## [1] 0.03944702
```

What is interesting is that the cloglog model does statistically as well as the estimated link parameter model while the logit does statistically worse in the second series but not the first. As such, while it wasn't expressly stated I am going to use my complementary log-log model estimates for the rest of the assignment.

2

Whether there is a noticable difference in the populations of the two series can be tested by running a model with the two series pooled (but not simply adding up the two populations. Keep them as two seperate observations at each doseage). The logliklihood of this unified model can be compared to the sum of the two logliklihood's from the seperate model. This will be a 3 degrees of freedom difference in the models. Note unified is used to represent the combined data sets.

```
## $par
## [1] -39.443876   9.541716
##
## $value
## [1] 25.16564
##
## $counts
## function gradient
##       83       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##          [,1]      [,2]
## [1,] 157.5938  652.9903
## [2,] 652.9903 2707.4488
```

```
chi_unified <- -2*(clogmod2$value + clogmod1$value - cloglog_unified$value)
pchisq(chi_unified, 3, lower.tail = FALSE)
```

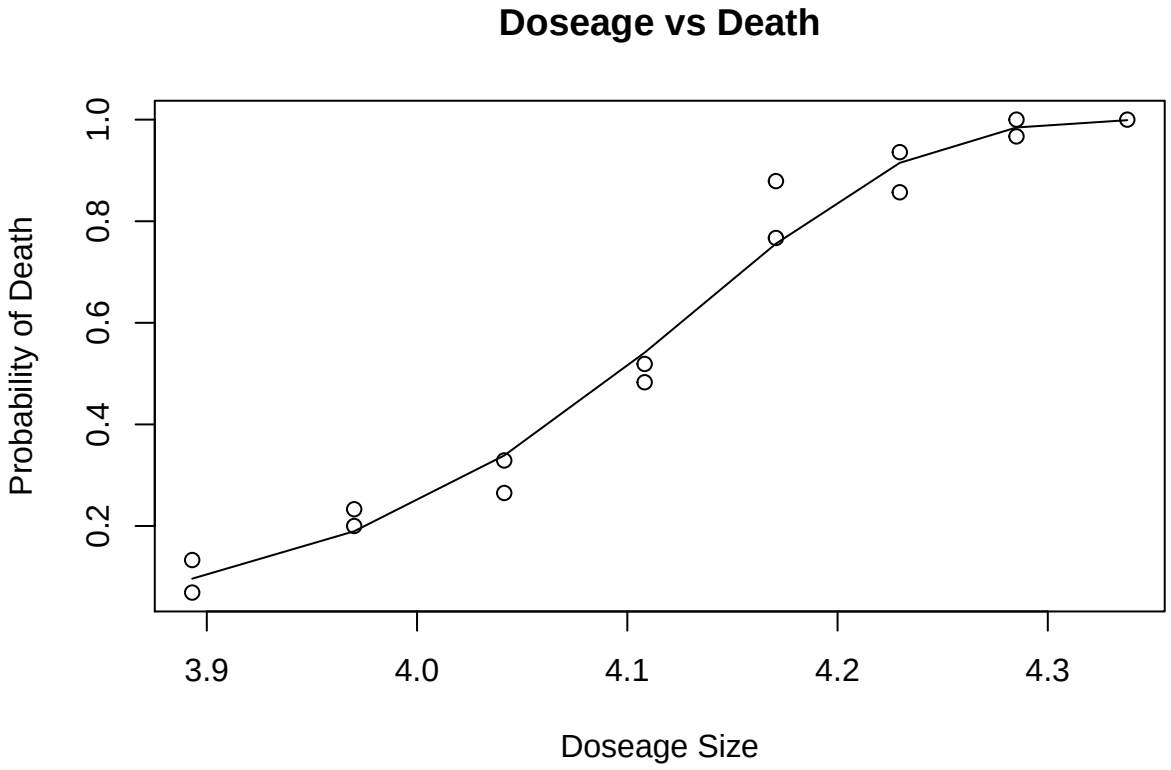
```
## [1] 0.9960688
```

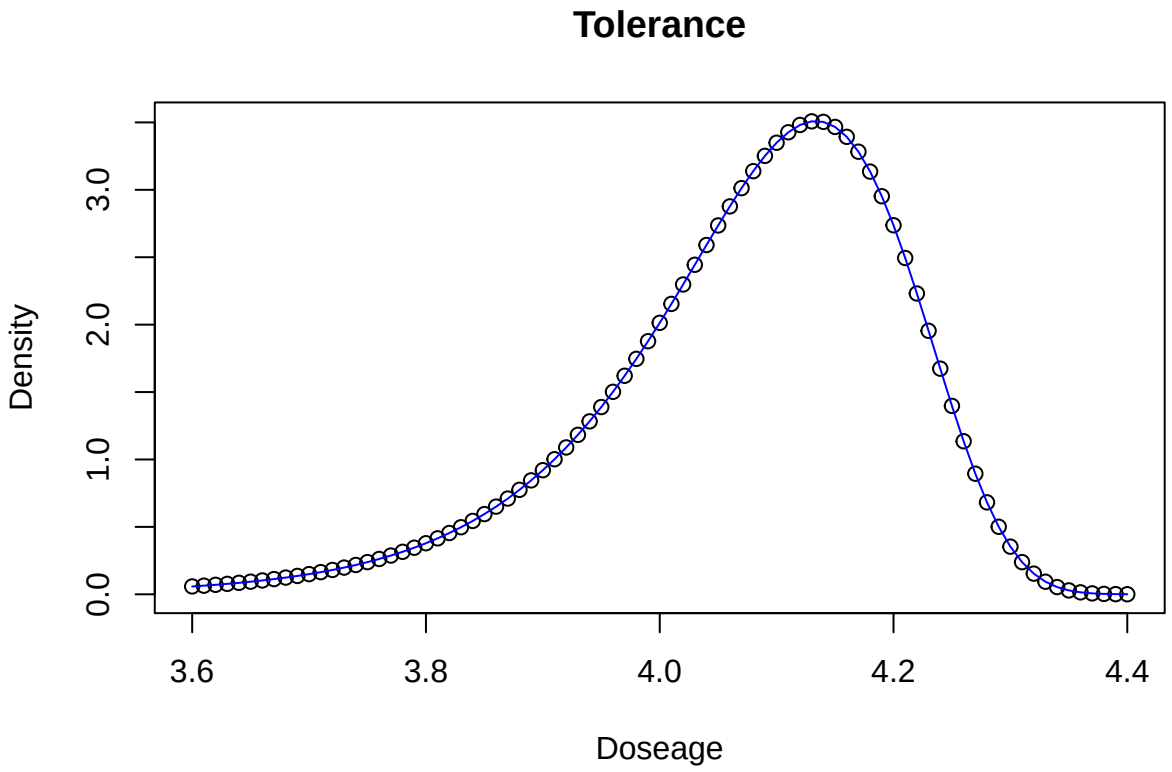
So then we find not statistical difference between running two series and just running one unified data set.

3

The final model then is a complementary log-log linked binomial. Please see the model presented in question 2 to see the estimates

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



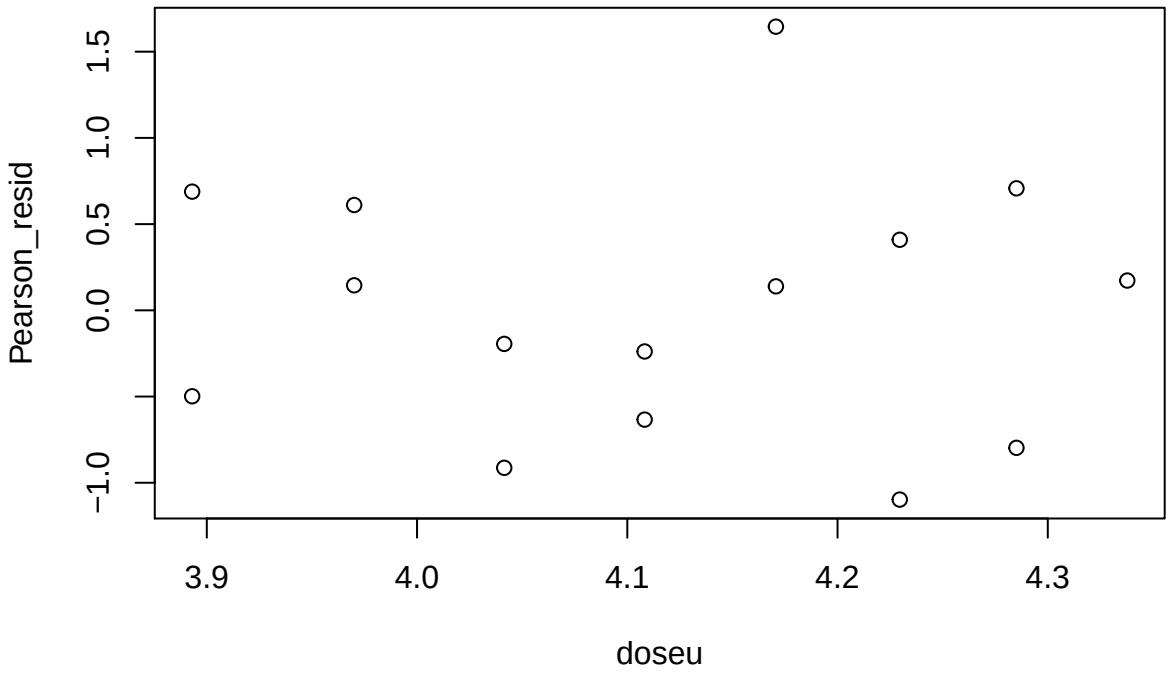


4 Assessment

The Pearson residuals were taken from the data and plotted. There appears to be one possible outlier and maybe a general U-shape in the data.

After that a Pearson Residual Chi Square test was ran to see if our model was statstically worse than claiming the doseages are in fact multinomial (the exhuastive model). Our p-value of .4647 suggests that there is no discernable difference in the two proposed models. It should be noted that the Pearson Residual Chi Square test is somewhat unstable for cells that have counts with less than 5 successes (e.g. the lower end doseages in this case).

```
## The following objects are masked _by_ .GlobalEnv:
##
##     doseu, nu, pu
```




```
## [1] 0.4647649
```

5

Computations for this assignment were mostly hand made. Avaliable on github at vinny-paris/One_Parameter_Pregibon_Link if you would like an inside look. The liklihood derivation is attached. Both the logit and the cloglog models were estimated withen my likelihood function with lambda = 1 for logit and lambda = .00001 for cloglog (lambda needs to approach 0).