Week 05 Homework

Brian Ritz

Friday, February 13, 2015

** POISSON REGRESSION AND NEGATIVE BINOMIAL REGRESSION **

Import the file:

```r
measles.data<-read.csv(file="MeaslesImmunizationCoverageAndOutbreaks.csv")
measles.data<-as.data.frame(measles.data)
```

We will fit the poisson model– the link function for a poisson model is the log function.

```
##  [1] "coefficients"      "residuals"         "fitted.values"
##  [4] "effects"           "R"                 "rank"
##  [7] "qr"                "family"            "linear.predictors"
## [10] "deviance"          "aic"               "null.deviance"
## [13] "iter"              "weights"           "prior.weights"
## [16] "df.residual"       "df.null"           "y"
## [19] "converged"         "boundary"          "model"
## [22] "call"              "formula"           "terms"
## [25] "data"              "offset"            "control"
## [28] "method"            "contrasts"         "xlevels"
```

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

```
## glm(formula = Outbreaks ~ Coverage, family = poisson, data = measles.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5168  -0.5378   0.0102   0.6222   1.3876
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)     5.512      0.175   31.56  < 2e-16 ***
## Coverage       -0.783      0.196   -3.99  6.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 52.549  on 53  degrees of freedom
## Residual deviance: 37.549  on 52  degrees of freedom
## AIC: 400.7
##
## Number of Fisher Scoring iterations: 3
```

We will get the linear predictors from the model object:

```
measles.poisson.model$linear.predictors
```

```
##      1     2     3     4     5     6     7     8     9    10    11    12
## 4.807 4.815 4.821 4.823 4.828 4.810 4.853 4.807 4.774 4.767 4.799 4.799
##     13    14    15    16    17    18    19    20    21    22    23    24
```

```
## 4.787 4.817 4.809 4.795 4.785 4.812 4.835 4.836 4.810 4.774 4.763 4.834
##    25    26    27    28    29    30    31    32    33    34    35    36
## 4.807 4.803 4.816 4.846 4.796 4.818 4.768 4.790 4.831 4.783 4.762 4.821
##    37    38    39    40    41    42    43    44    45    46    47    48
## 4.852 4.829 4.833 4.787 4.763 4.823 4.796 4.820 4.795 4.791 4.814 4.849
##    49    50    51    52    53    54
## 4.810 4.879 4.802 4.817 4.872 5.080
```

The linear predictors are the log of the fitted values. Below, we compare the outbreaks with the fitted values.

```
rbind(measles.data$Outbreaks,measles.poisson.model$fitted.values)
```

```
##             1     2   3     4   5     6     7     8     9    10    11    12
## [1,] 124.0 114.0 129 122.0 131 135.0 143.0 118.0 115.0 123.0 135.0 124.0
## [2,] 122.3 123.4 124 124.3 125 122.8 128.2 122.3 118.3 117.6 121.3 121.3
##          13    14    15    16    17  18    19  20    21    22    23    24
## [1,] 110.0 126.0 120.0 107.0 128.0 117 120.0 124 135.0 114.0 128.0 130.0
## [2,] 119.9 123.7 122.6 120.9 119.7 123 125.8 126 122.7 118.3 117.1 125.7
##          25    26    27    28    29    30    31    32    33    34  35    36
## [1,] 125.0 134.0 131.0 128.0 125.0 127.0 127.0 108.0 116.0 121.0 112 139.0
## [2,] 122.3 121.9 123.5 127.2 121.1 123.7 117.7 120.3 125.3 119.5 117 124.1
##        37    38    39    40    41    42    43    44    45    46    47    48
## [1,] 144 118.0 134.0  99.0 127.0 118.0 116.0  97.0 120.0 117.0 114.0 127.0
## [2,] 128 125.1 125.6 119.9 117.1 124.3 121.1 123.9 120.9 120.4 123.3 127.6
##          49    50    51    52    53    54
## [1,] 113.0 137.0 129.0 115.0 118.0 160.0
## [2,] 122.7 131.4 121.7 123.7 130.6 160.7
```

I will show that the link is logrithmic:

```
rbind(log(measles.poisson.model$fitted.values), measles.poisson.model$linear.predictors)
```

```
##              1     2     3     4     5    6     7     8     9    10    11
## [1,] 4.807 4.815 4.821 4.823 4.828 4.81 4.853 4.807 4.774 4.767 4.799
## [2,] 4.807 4.815 4.821 4.823 4.828 4.81 4.853 4.807 4.774 4.767 4.799
##          12    13    14    15    16    17    18    19    20   21    22
## [1,] 4.799 4.787 4.817 4.809 4.795 4.785 4.812 4.835 4.836 4.81 4.774
## [2,] 4.799 4.787 4.817 4.809 4.795 4.785 4.812 4.835 4.836 4.81 4.774
##          23    24    25    26    27    28    29    30    31   32    33
## [1,] 4.763 4.834 4.807 4.803 4.816 4.846 4.796 4.818 4.768 4.79 4.831
## [2,] 4.763 4.834 4.807 4.803 4.816 4.846 4.796 4.818 4.768 4.79 4.831
##          34    35    36    37    38    39    40    41    42    43   44
## [1,] 4.783 4.762 4.821 4.852 4.829 4.833 4.787 4.763 4.823 4.796 4.82
## [2,] 4.783 4.762 4.821 4.852 4.829 4.833 4.787 4.763 4.823 4.796 4.82
##          45    46    47    48   49    50    51    52    53   54
## [1,] 4.795 4.791 4.814 4.849 4.81 4.879 4.802 4.817 4.872 5.08
## [2,] 4.795 4.791 4.814 4.849 4.81 4.879 4.802 4.817 4.872 5.08
```

What would happen if the coverage changes by 1%?

```
## coverage change in percentage points
(1-exp(measles.poisson.model$coef[2]*.01))*100
```

```
## Coverage
##   0.7798
```

** NEGATIVE BINOMIAL REGRESSION **

```r
library(MASS)
neg.bin.regression.model <- glm.nb(Outbreaks ~ Coverage, data = measles.data, init.theta = 2934552)
```

```
## Warning: iteration limit reached
## Warning: iteration limit reached
```

```r
summary(neg.bin.regression.model)
```

```
##
## Call:
## glm.nb(formula = Outbreaks ~ Coverage, data = measles.data, init.theta = 2934442.091,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -2.5167   -0.5377   0.0102   0.6222   1.3875
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    5.512      0.175    31.56  < 2e-16 ***
## Coverage      -0.783      0.196    -3.99  6.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for Negative Binomial(2934442) family taken to be 1)
##
##     Null deviance: 52.547  on 53  degrees of freedom
## Residual deviance: 37.547  on 52  degrees of freedom
## AIC: 402.7
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  2934442
##           Std. Err.:  59522631
## Warning while fitting theta: iteration limit reached
##
##  2 x log-likelihood:  -396.7
```

The negative binomial regression also finds that coverage is significantly negatively correlated with the frequency of cases of Measles. The coefficient on coverage estimated by the negative binomial model is -0.7828 – the same as for the poisson model. The standard error of the poisson estimate for the coefficient on coverage is 0.1963, the negative binomial's standard error is the same–0.1963. The deviances are also roughly equal…just 0.002 difference.

I can conclude that the poisson model is appropriate for this data.