STAT 8020 R Lab 17: Poisson Regression II

Whitney

October 28, 2020

Contents

Species Diversity on the Galapagos Islands

```r
library(faraway)
data(gala)
gala <- gala[, -2] # remove "Endemics"
gala
```

```
##              Species     Area Elevation Nearest  Scruz Adjacent
## Baltra            58    25.09       346     0.6    0.6     1.84
## Bartolome         31     1.24       109     0.6   26.3   572.33
## Caldwell           3     0.21       114     2.8   58.7     0.78
## Champion          25     0.10        46     1.9   47.4     0.18
## Coamano            2     0.05        77     1.9    1.9   903.82
## Daphne.Major      18     0.34       119     8.0    8.0     1.84
## Daphne.Minor      24     0.08        93     6.0   12.0     0.34
## Darwin            10     2.33       168    34.1  290.2     2.85
## Eden               8     0.03        71     0.4    0.4    17.95
## Enderby            2     0.18       112     2.6   50.2     0.10
## Espanola          97    58.27       198     1.1   88.3     0.57
## Fernandina        93   634.49      1494     4.3   95.3  4669.32
## Gardner1          58     0.57        49     1.1   93.1    58.27
## Gardner2           5     0.78       227     4.6   62.2     0.21
## Genovesa          40    17.35        76    47.4   92.2   129.49
## Isabela          347  4669.32      1707     0.7   28.1   634.49
## Marchena          51   129.49       343    29.1   85.9    59.56
## Onslow             2     0.01        25     3.3   45.9     0.10
## Pinta            104    59.56       777    29.1  119.6   129.49
## Pinzon           108    17.95       458    10.7   10.7     0.03
## Las.Plazas        12     0.23        94     0.5    0.6    25.09
## Rabida            70     4.89       367     4.4   24.4   572.33
## SanCristobal     280   551.62       716    45.2   66.6     0.57
## SanSalvador      237   572.33       906     0.2   19.8     4.89
## SantaCruz        444   903.82       864     0.6    0.0     0.52
## SantaFe           62    24.08       259    16.5   16.5     0.52
## SantaMaria       285   170.92       640     2.6   49.2     0.10
## Seymour           44     1.84       147     0.6    9.6    25.09
## Tortuga           16     1.24       186     6.8   50.9    17.95
## Wolf              21     2.85       253    34.1  254.7     2.33
```

```
lmFit <- lm(Species ~ ., gala)
summary(lmFit)
```

```
##
## Call:
## lm(formula = Species ~ ., data = gala)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -111.679  -34.898   -7.862   33.460  182.584
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.068221  19.154198   0.369 0.715351
## Area        -0.023938   0.022422  -1.068 0.296318
## Elevation    0.319465   0.053663   5.953 3.82e-06 ***
## Nearest      0.009144   1.054136   0.009 0.993151
## Scruz       -0.240524   0.215402  -1.117 0.275208
## Adjacent    -0.074805   0.017700  -4.226 0.000297 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.98 on 24 degrees of freedom
## Multiple R-squared:  0.7658, Adjusted R-squared:  0.7171
## F-statistic:  15.7 on 5 and 24 DF,  p-value: 6.838e-07
```

```
PoiFit <- glm(Species ~ ., family = poisson, gala)
summary(PoiFit)
```

```
##
## Call:
## glm(formula = Species ~ ., family = poisson, data = gala)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -8.2752  -4.4966  -0.9443   1.9168  10.1849
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.155e+00  5.175e-02  60.963  < 2e-16 ***
## Area        -5.799e-04  2.627e-05 -22.074  < 2e-16 ***
## Elevation    3.541e-03  8.741e-05  40.507  < 2e-16 ***
## Nearest      8.826e-03  1.821e-03   4.846 1.26e-06 ***
## Scruz       -5.709e-03  6.256e-04  -9.126  < 2e-16 ***
## Adjacent    -6.630e-04  2.933e-05 -22.608  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 3510.73  on 29  degrees of freedom
## Residual deviance: 716.85  on 24  degrees of freedom
## AIC: 889.68
##
```

```
## Number of Fisher Scoring iterations: 5
```

```r
step(PoiFit)
```

```
## Start:  AIC=889.68
## Species ~ Area + Elevation + Nearest + Scruz + Adjacent
##
##             Df Deviance    AIC
## <none>           716.85  889.68
## - Nearest   1   739.41  910.24
## - Scruz     1   813.62  984.45
## - Area      1  1204.35 1375.18
## - Adjacent  1  1341.45 1512.29
## - Elevation 1  2389.57 2560.40

##
## Call:  glm(formula = Species ~ Area + Elevation + Nearest + Scruz +
##     Adjacent, family = poisson, data = gala)
##
## Coefficients:
## (Intercept)         Area    Elevation      Nearest        Scruz     Adjacent
##   3.1548079   -0.0005799    0.0035406    0.0088256   -0.0057094   -0.0006630
##
## Degrees of Freedom: 29 Total (i.e. Null);  24 Residual
## Null Deviance:        3511
## Residual Deviance: 716.8      AIC: 889.7
```

Wafer Quality

```r
Freq <- c(320,14,80,36)
Particle <- gl(2, 1, 4, labels = c("No", "Yes"))
Quality <- gl(2, 2, labels = c("Good", "Bad"))
(wafer <- data.frame(Freq, Particle, Quality))
```

```
##   Freq Particle Quality
## 1  320       No    Good
## 2   14      Yes    Good
## 3   80       No     Bad
## 4   36      Yes     Bad
```

```r
xtabs(Freq ~ Quality + Particle)
```

```
##         Particle
## Quality  No Yes
##    Good 320  14
##    Bad   80  36
```

Poisson Model

```r
mod1 <- glm(Freq ~ Quality + Particle, family = poisson)
sumary(mod1)
```

```
##             Estimate Std. Error  z value  Pr(>|z|)
## (Intercept)  5.69336    0.05720  99.5350 < 2.2e-16
## QualityBad  -1.05755    0.10777  -9.8129 < 2.2e-16
## ParticleYes -2.07944    0.15000 -13.8630 < 2.2e-16
```

3

```
##
## n = 4 p = 3
## Deviance = 54.03045 Null Deviance = 474.09877 (Difference = 420.06832)
```

```r
drop1(mod1, test = "Chi")
```

```
## Single term deletions
##
## Model:
## Freq ~ Quality + Particle
##          Df Deviance    AIC    LRT  Pr(>Chi)
## <none>        54.03  83.77
## Quality   1   164.22 191.96 110.19 < 2.2e-16 ***
## Particle  1   363.91 391.66 309.88 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multinomial Model

```r
n = 450
(pp <- prop.table(xtabs(Freq ~ Particle)))
```

```
## Particle
##        No       Yes
## 0.8888889 0.1111111
```

```r
(qp <- prop.table(xtabs(Freq ~ Quality)))
```

```
## Quality
##      Good       Bad
## 0.7422222 0.2577778
```

```r
(exp <- outer(qp, pp) * n)
```

```
##         Particle
## Quality       No      Yes
##    Good 296.8889 37.11111
##    Bad  103.1111 12.88889
```

```r
(obs <- xtabs(Freq ~ Quality + Particle))
```

```
##         Particle
## Quality  No Yes
##    Good 320  14
##    Bad   80  36
```

```r
(2 * sum(obs * log(obs / exp)))
```

```
## [1] 54.03045
```

Binomial Model

```r
(m <- matrix(Freq, nrow = 2))
```

```
##      [,1] [,2]
## [1,]  320   80
## [2,]   14   36
```

```r
(binFit <- glm(m ~ 1, family = binomial))
```

```
##
## Call:  glm(formula = m ~ 1, family = binomial)
##
## Coefficients:
## (Intercept)
##        1.058
##
## Degrees of Freedom: 1 Total (i.e. Null);  1 Residual
## Null Deviance:        54.03
## Residual Deviance: 54.03     AIC: 66.19
```

```r
predict(binFit, type = "response")
```

```
##         1         2
## 0.7422222 0.7422222
```

Hypergeometric Model

```r
fisher.test(obs)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  obs
## p-value = 2.955e-13
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    5.090628 21.544071
## sample estimates:
## odds ratio
##    10.21331
```