

Lab 13 – Generalized Linear Models

November 12 & 13, 2018
FANR 6750

Richard Chandler and Bob Cooper

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots$$

$$y_i \sim \text{Binomial}(N, p_i)$$

where:

N is the number of “trials” (e.g. coin flips)

p_i is the probability of a success for sample unit i

PRESENCE-ABSENCE AND ABUNDANCE

```
frogData[1:25,] # First 25 rows

##   presence abundance elevation habitat
## 1         0         0         58      Oak
## 2         1         7        191      Oak
## 3         0         0         43      Oak
## 4         1        11        374      Oak
## 5         1        11        337      Oak
## 6         1         1         64      Oak
## 7         1         4        195      Oak
## 8         1         6        263      Oak
## 9         0         0        181      Oak
## 10        1         1         59      Oak
## 11        1        50        489      Maple
## 12        1         5        317      Maple
## 13        0         0         12      Maple
## 14        1         4        245      Maple
## 15        1        47        474      Maple
## 16        1         1         83      Maple
## 17        1        46        467      Maple
## 18        1        51        485      Maple
## 19        1        23        335      Maple
## 20        0         0         20      Maple
## 21        1        27        430      Pine
## 22        1         2        223      Pine
## 23        0         0         68      Pine
## 24        1        50        483      Pine
## 25        0         0         78      Pine
```

LOGISTIC REGRESSION USING glm

```
fm1 <- glm(presence ~ habitat + elevation,
           family=binomial(link="logit"), data=frogData)
```

```
summary(fm1)

##
## Call:
## glm(formula = presence ~ habitat + elevation, family = binomial(link = "logit"),
##      data = frogData)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8843  -0.6169   0.1674   0.6050   1.3775
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.092759   1.055124  -1.036   0.3004
## habitatMaple  0.096781   1.367518   0.071   0.9436
## habitatPine  -0.240443   1.154650  -0.208   0.8350
## elevation     0.013658   0.006011   2.272   0.0231 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 34.795  on 29  degrees of freedom
## Residual deviance: 23.132  on 26  degrees of freedom
## AIC: 31.132
##
## Number of Fisher Scoring iterations: 6
```

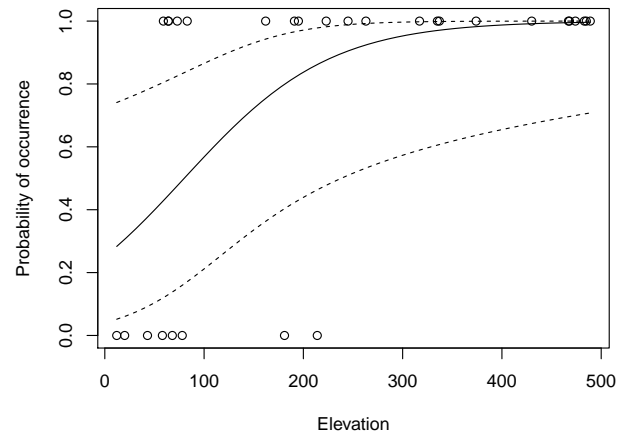
```
predData.elev <- data.frame(elevation=seq(12, 489, length=50),
                           habitat="Oak")
head(predData.elev)
```

```
## elevation habitat
## 1 12.00000 Oak
## 2 21.73469 Oak
## 3 31.46939 Oak
## 4 41.20408 Oak
## 5 50.93878 Oak
## 6 60.67347 Oak
```

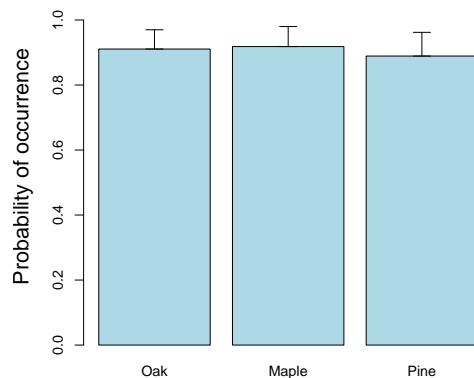
To get confidence intervals on (0,1) scale, predict on logit (link) scale and then backtransform using inverse-link

```
pred.link <- predict(fm1, newdata=predData.elev, se.fit=TRUE, type="link")
predData.elev$p <- plogis(pred.link$fit)
predData.elev$lower <- plogis(pred.link$fit - 1.96*pred.link$se.fit)
predData.elev$upper <- plogis(pred.link$fit + 1.96*pred.link$se.fit)
```

```
plot(p ~ elevation, data=predData.elev, type="l", ylim=c(0,1),
     xlab="Elevation", ylab="Probability of occurrence")
points(presence ~ elevation, frogData)
lines(lower ~ elevation, data=predData.elev, lty=2)
lines(upper ~ elevation, data=predData.elev, lty=2)
```



```
predData.hab <- data.frame(habitat=c("Oak", "Maple", "Pine"), elevation=250)
pred <- predict(fm1, newdata=predData.hab, se.fit=TRUE, type="link")
bp <- barplot(plogis(pred$fit), ylab="Probability of occurrence", cex.lab=1.5,
              names=c("Oak", "Maple", "Pine"), col="lightblue", ylim=c(0, 1.1))
arrows(bp, plogis(pred$fit), bp, plogis(pred$fit + pred$se.fit),
       angle=90, code=3, length=0.1)
```



$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots$$

$$y_i \sim \text{Poisson}(\lambda_i)$$

where:

λ_i is the expected value of y_i

```
fm2 <- glm(abundance ~ habitat + elevation,
           family=poisson(link="log"), data=frogData)
```

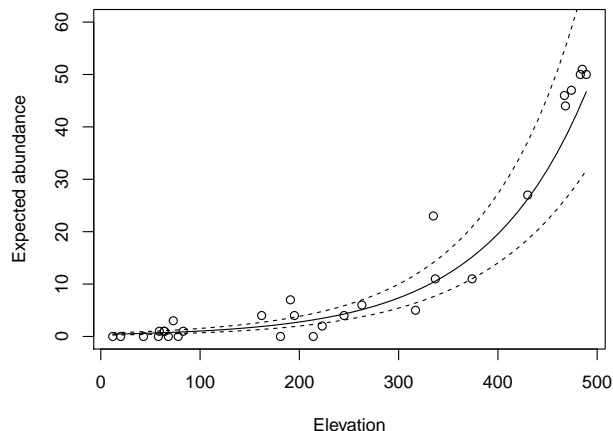
```
summary(fm2)
```

```
##
## Call:
## glm(formula = abundance ~ habitat + elevation, family = poisson(link = "log"),
##      data = frogData)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6308  -1.0810  -0.1067   0.3353   2.7935
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.9403442  0.2430730  -3.869 0.000109 ***
## habitatMaple  0.1533915  0.1971009   0.778 0.436428
## habitatPine   0.0881110  0.1994981   0.442 0.658733
## elevation     0.0097836  0.0006291  15.551 < 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: 700.762  on 29  degrees of freedom
## Residual deviance:  44.891  on 26  degrees of freedom
## AIC: 140.13
##
## Number of Fisher Scoring iterations: 5
```

To get confidence intervals on $(0, \infty)$ scale, predict on log (link) scale and then backtransform using inverse-link

```
pred.link <- predict(fm2, newdata=predData.elev, se.fit=TRUE, type="link")
predData.elev$lambda <- exp(pred.link$fit) # exp is the inverse-link function
predData.elev$lower <- exp(pred.link$fit - 1.96*pred.link$se.fit)
predData.elev$upper <- exp(pred.link$fit + 1.96*pred.link$se.fit)
```

```
plot(lambda ~ elevation, predData.elev, type="l", ylim=c(0,60),
     xlab="Elevation", ylab="Expected abundance")
points(abundance ~ elevation, frogData)
lines(lower ~ elevation, predData.elev, lty=2)
lines(upper ~ elevation, predData.elev, lty=2)
```



Researchers want to know how latitude and landscape type influence the probability that American Crows are infected by West Nile Virus. One hundred crows are captured and tested for West Nile Virus in urban and rural landscapes spanning a latitude gradient.

- (1) Fit a logistic regression model to the `crowData.csv` dataset to assess the effects of latitude and landscape type
- (2) Interpret the parameter estimates
- (3) Plot the relationship between infection probability and latitude, for rural and urban landscapes, on the same graph
- (4) Include the data points (color coded by landscape) and a legend in the graph
- (5) Include confidence intervals